Journal of Multivariate Analysis **82**, 299–330 (2002)
doi:10.1006/jmva.2001.2034

Exact Misclassification Probabilities for Plug-In Normal Quadratic Discriminant Functions

II. The Heterogeneous Case

H. Richard McFarland, III

Digitas, Inc.

and

Donald St. P. Richards[1]

Institute for Advanced Study and University of Virginia

Received May 18, 1999; published online February 5, 2002

We consider the problem of discriminating between two independent multivariate normal populations, $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, having distinct mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and distinct covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$. The parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_1$, and $\boldsymbol{\Sigma}_2$ are unknown and are estimated by means of independent random training samples from each population. We derive a stochastic representation for the exact distribution of the "plug-in" quadratic discriminant function for classifying a new observation between the two populations. The stochastic representation involves only the classical standard normal, chi-square, and F distributions and is easily implemented for simulation purposes. Using Monte Carlo simulation of the stochastic representation we provide applications to the estimation of misclassification probabilities for the well-known iris data studied by Fisher (*Ann. Eugen.* **7** (1936), 179–188); a data set on corporate financial ratios provided by Johnson and Wichern (*Applied Multivariate Statistical Analysis*, 4th ed., Prentice–Hall, Englewood Cliffs, NJ, 1998); and a data set analyzed by Reaven and Miller (*Diabetologia* **16** (1979), 17–24) in a classification of diabetic status.

AMS 1991 subject classifications: 62H10; 62H30; 62E15.

Key words and phrases: apparent error rate; Bessel function of matrix argument; corporate financial data; cross-validation; diabetes data; discriminant analysis; hold-out method; iris data; misclassification probability; multivariate gamma function; multivariate normal distribution; resubstitution method; stochastic representation; Wishart distribution.

[1] Supported in part by a grant to the Institute for Advanced Study from The Bell Fund, and by the National Science Foundation under grant DMS-9703705.

0047-259X/02 $35.00

1. INTRODUCTION

Let $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ denote $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, respectively, two independent multivariate normal populations with mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$. In discriminant analysis a classical problem is to study procedures for classifying an observation $\mathbf{y}$ into one of the populations $\mathbf{\Pi}_1$ or $\mathbf{\Pi}_2$, and special attention has been paid to those procedures which minimize the total expected cost of misclassification. Comprehensive accounts of the theory and applications of this problem are provided by Anderson [1], Johnson and Wichern [8], McLachlan [15], and Muirhead [16].

For the case in which the parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_1$, and $\boldsymbol{\Sigma}_2$ are known, the discriminant rule resulting from the procedure which minimizes the total expected cost of misclassification is a quadratic function of the observation vector $\mathbf{y}$, viz.,

$$Q = \frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}_2)' \boldsymbol{\Sigma}_2^{-1}(\mathbf{y}-\boldsymbol{\mu}_2) - \frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}_1)' \boldsymbol{\Sigma}_1^{-1}(\mathbf{y}-\boldsymbol{\mu}_1) + \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|}. \tag{1.1}$$

In this case, descriptions of the distribution of the discriminant function, the best regions of classification and the corresponding probabilities of misclassification are classical results; cf. [1, 8, 15, 16].

In practice, the parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_1$, and $\boldsymbol{\Sigma}_2$ are unknown. Then we estimate these parameters by means of independent random "training samples," $\mathbf{y}_1^{(1)}, ..., \mathbf{y}_{N_1}^{(1)}$ from $\mathbf{\Pi}_1$ and $\mathbf{y}_1^{(2)}, ..., \mathbf{y}_{N_2}^{(2)}$ from $\mathbf{\Pi}_2$, where $N_1, N_2 > p$. Suppose, for instance, that $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ have a common covariance matrix, denoted $\boldsymbol{\Sigma}$; then the log-term in (1.1) is identically zero, the quadratic terms in $\mathbf{y}$ cancel each other, and we obtain a discriminant function which is a linear function of $\mathbf{y}$. We estimate $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ by the sample means $\bar{\mathbf{y}}_1 = \sum_{i=1}^{N_1} \mathbf{y}_i^{(1)}/N_1$ and $\bar{\mathbf{y}}_2 = \sum_{i=1}^{N_2} \mathbf{y}_i^{(2)}/N_2$, respectively; and we estimate $\boldsymbol{\Sigma}$ using the pooled covariance matrix $\mathbf{S} = ((N_1-1)\,\mathbf{S}_1 + (N_2-1)\,\mathbf{S}_2)/(N_1+N_2-2)$ where, for $g = 1, 2$,

$$\mathbf{S}_g = \frac{1}{N_g - 1}\sum_{i=1}^{N_g} (\mathbf{y}_i^{(g)} - \bar{\mathbf{y}}_g)(\mathbf{y}_i^{(g)} - \bar{\mathbf{y}}_g)' \tag{1.2}$$

is the sample covariance matrix corresponding to the gth sample. Substituting these estimates into the resulting likelihood ratio yields a "plug-in" discriminant function, commonly denoted by W, which has been studied by several authors; cf. [1, Sect. 6.5]. In particular, we note that stochastic representations for the exact distribution of W were derived by Bowker [5].

Another problem of great interest arises in the equal-means case in which the populations $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ are assumed to have a common mean vector $\boldsymbol{\mu}$, and distinct covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$. This problem was studied first by Okamoto [17], and later by Bartlett and Please [3] who gave an application of classification based on a data set collected by Stocks [20]. After drawing independent random training samples, $\mathbf{y}_1^{(1)}, \ldots, \mathbf{y}_{N_1}^{(1)}$ from $\boldsymbol{\Pi}_1$ and $\mathbf{y}_1^{(2)}, \ldots, \mathbf{y}_{N_2}^{(2)}$ from $\boldsymbol{\Pi}_2$, a plug-in quadratic discriminant function is obtained from (1.1) by substituting for $\boldsymbol{\mu}$ the pooled sample mean $\bar{\mathbf{y}} = (N_1 \bar{\mathbf{y}}_1 + N_2 \bar{\mathbf{y}}_2)/(N_1 + N_2)$, and substituting for $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ the sample covariance matrices $\mathbf{S}_1$ and $\mathbf{S}_2$, respectively. A stochastic representation for the exact distribution of the corresponding plug-in discriminant function was obtained recently by McFarland [13] and McFarland and Richards [14]. The techniques utilized in [13, 14] made crucial use of certain Bessel functions of matrix argument of the second kind, defined by Herz [9].

Let us return to the general situation in which the mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are unknown and distinct, and also the covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are unknown and distinct. Then the *plug-in quadratic discriminant function* corresponding to (1.1), i.e.,

$$\hat{Q} = \frac{1}{2} (\mathbf{y} - \bar{\mathbf{y}}_2)' \, \mathbf{S}_2^{-1} (\mathbf{y} - \bar{\mathbf{y}}_2) - \frac{1}{2} (\mathbf{y} - \bar{\mathbf{y}}_1)' \, \mathbf{S}_1^{-1} (\mathbf{y} - \bar{\mathbf{y}}_1) + \frac{1}{2} \log \frac{|\mathbf{S}_2|}{|\mathbf{S}_1|}, \tag{1.3}$$

is obtained by replacing each unknown parameter by its unbiased estimate.

As we noted in [14], in many applications of normal discriminant analysis, the sample sizes N_1 and N_2 are not large enough for asymptotic approximations of quadratic discriminant functions to be expected to yield accurate estimates of the probabilities of misclassification. In this paper we develop exact procedures for estimating the misclassification probabilities associated with the plug-in quadratic discriminant function $\hat{Q}$. These results are obtained as a consequence of a stochastic representation for the exact distribution of $\hat{Q}$. As in the article [14] our approach is based on a decomposition of the characteristic function of $\hat{Q}$, and our results again utilize the Bessel functions of matrix argument of the second kind. Having derived the exact stochastic representations, we use Monte Carlo simulation to estimate the corresponding probabilities of misclassification.

In Section 2 we state the main results on the stochastic representation of $\hat{Q}$. In Section 3 we apply the stochastic representation to estimate misclassification probabilities for the well-known iris data studied by R. A. Fisher [7], data on corporate financial ratios provided by Johnson and Wichern [8], and data utilized by Reaven and Miller [18] to study the classification of diabetic status from chemical decomposition of blood samples. For each of these three examples, we compare the estimates derived using the stochastic representations with estimates obtained through the resubstitution and holdout methods. In Section 4 we list some preliminary details necessary for the proof of the main result.

Finally, in Section 5 we provide the proof of the main result, Theorem 2.1. We remark that the proof is based on techniques similar to those used in [14]. However, the technical details needed in the present paper, and especially the application of Lemma 4.6 and the formulas (5.7)–(5.17), are significantly more complicated than in [14].

2. THE STOCHASTIC REPRESENTATION FOR $\hat{Q}$

For simplicity, we will assume throughout that the prior probabilities for $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ are equal and also that the costs of misclassification are equal (cf. Johnson and Wichern [8, pp. 634–637]). We also use the standard notation $P(2 \mid 1)$ for the probability of misclassifying an observation $\mathbf{y}$ into $\mathbf{\Pi}_2$ when, in fact, $\mathbf{y} \in \mathbf{\Pi}_1$; and we use $P(1 \mid 2)$ to denote the probability of misclassifying $\mathbf{y}$ into $\mathbf{\Pi}_1$ when, in fact, $\mathbf{y} \in \mathbf{\Pi}_2$. Specifically, we have

$$P(2 \mid 1) = \mathrm{P}\{\hat{Q} \leqslant 0 \mid \mathbf{y} \in \mathbf{\Pi}_1\}, \qquad P(1 \mid 2) = \mathrm{P}\{\hat{Q} > 0 \mid \mathbf{y} \in \mathbf{\Pi}_2\}.$$

Since the costs of misclassification are equal then, TPM, the total expected cost of misclassification is

$$\mathrm{TPM} = \tfrac{1}{2}\,(P(1 \mid 2) + P(2 \mid 1)).$$

Let $\mathbf{H}$ be a $p \times p$ orthogonal matrix which diagonalizes the matrix $\mathbf{\Sigma}_2^{-1/2} \mathbf{\Sigma}_1 \mathbf{\Sigma}_2^{-1/2}$. We will write

$$\mathbf{\Sigma}_2^{-1/2} \mathbf{\Sigma}_1 \mathbf{\Sigma}_2^{-1/2} = \mathbf{H} \mathbf{\Lambda} \mathbf{H}', \tag{2.1}$$

where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, ..., \lambda_p)$, a diagonal matrix. Clearly, the entries $\lambda_1, ..., \lambda_p$ on the main diagonal are the eigenvalues of $\mathbf{\Sigma}_2^{-1} \mathbf{\Sigma}_1$.

Define the column vector

$$\boldsymbol{\mu} := \mathbf{H}' \mathbf{\Sigma}_2^{-1/2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \tag{2.2}$$

and denote its coordinates by $\mu_1, ..., \mu_p$. For $j = 1, ..., p$, set

$$\omega_{3j} = \left(\frac{N_1 N_2 \lambda_j}{(N_1 + 1)(\lambda_j N_2 + 1)} \right)^{1/2} \tag{2.3}$$

and

$$\gamma_j = \left(\lambda_j + \frac{1}{N_2} \right)^{-1/2} \mu_j. \tag{2.4}$$

The stochastic representation for $\hat{Q}$ will involve mutually independent random variables $T_1, T_2, Z_{11}, \ldots, Z_{1p}, Z_{21}, \ldots, Z_{2p}$, and $F_1, \ldots, F_{p-1}$, where for $g = 1, 2$, $T_g \stackrel{d}{=} \chi^2_{N_g - p}$, a chi-square distribution with $N_g - p$ degrees of freedom; for $j = 1, \ldots, p$, $Z_{gj} \stackrel{d}{=} N(0, 1)$, a standard normal random variable; and for $j = 1, \ldots, p-1$, F_j is F-distributed with $(N_2 - j, N_1 - j)$ degrees of freedom. For $g = 1, 2$ we set $n_g := N_g - 1$; and for $j = 1, \ldots, p$ we define

$$v_1 \stackrel{d}{=} \frac{n_1(N_1+1)}{N_1 T_1}, \qquad v_{2j} \stackrel{d}{=} \frac{n_2(\lambda_j + N_2^{-1})}{T_2}. \tag{2.5}$$

The entities in Eqs. (2.3)–(2.5) arise in the stochastic representation for $\hat{Q}$ when $\mathbf{y} \in \mathbf{\Pi}_1$.

To study the case in which $\mathbf{y} \in \mathbf{\Pi}_2$ we define, for $j = 1, \ldots, p$,

$$\tilde{\omega}_{3j} = \left(\frac{N_1 N_2}{(N_2+1)(\lambda_j + N_1)}\right)^{1/2}, \tag{2.6}$$

$$\tilde{\gamma}_j = \left(1 + \frac{\lambda_j}{N_1}\right)^{-1/2} \mu_j, \tag{2.7}$$

and

$$\tilde{v}_1 \stackrel{d}{=} \frac{n_2(N_2+1)}{N_2 T_2}, \qquad \tilde{v}_{2j} \stackrel{d}{=} \frac{n_1(\lambda_j^{-1} + N_1^{-1})}{T_1}. \tag{2.8}$$

With these notations in place, we can now state the main result.

THEOREM 2.1. *If* $\mathbf{y} \in \mathbf{\Pi}_1$ *then the plug-in quadratic discriminant function* $\hat{Q}$ *in* (1.3) *satisfies the stochastic representation*

$$\begin{aligned}\hat{Q} \stackrel{d}{=} & \frac{1}{2} \sum_{j=1}^{p} \left[v_{2j}(\omega_{3j} Z_{1j} + (1-\omega_{3j}^2)^{1/2} Z_{2j} + \gamma_j)^2 - v_1 Z_{1j}^2\right] \\ & + \frac{1}{2}\left[\log\left(\frac{n_1^p T_2}{n_2^p T_1}\right) - \log|\Sigma_2^{-1}\Sigma_1| + \sum_{j=1}^{p-1} \log\left(\frac{N_2 - j}{N_1 - j} F_j\right)\right].\end{aligned} \tag{2.9}$$

Further, if $\mathbf{y} \in \mathbf{\Pi}_2$ *then*

$$\begin{aligned}\hat{Q} \stackrel{d}{=} & \frac{1}{2} \sum_{j=1}^{p} \left[\tilde{v}_1 Z_{1j}^2 - \tilde{v}_{2j}(\tilde{\omega}_{3j} Z_{1j} + (1-\tilde{\omega}_{3j}^2)^{1/2} Z_{2j} + \tilde{\gamma}_j)^2\right] \\ & + \frac{1}{2}\left[\log\left(\frac{n_1^p T_2}{n_2^p T_1}\right) - \log|\Sigma_2^{-1}\Sigma_1| + \sum_{j=1}^{p-1} \log\left(\frac{N_2 - j}{N_1 - j} F_j\right)\right].\end{aligned} \tag{2.10}$$

For the case in which the training samples are of equal size, Theorem 2.1 reduces to the following result.

COROLLARY 2.2. *Suppose $N_1 = N_2 \equiv N$. If $\mathbf{y} \in \mathbf{\Pi}_1$ then*

$$\hat{Q} \stackrel{d}{=} \frac{N-1}{2N} \sum_{j=1}^{p} \left[\frac{(\lambda_j N+1)}{\mathrm{T}_2} (\omega_{3j} \mathrm{Z}_{1j} + (1-\omega_{3j}^2)^{-1/2} \mathrm{Z}_{2j} + \gamma_j)^2 - \frac{(N+1)}{\mathrm{T}_1} \mathrm{Z}_{1j}^2 \right]$$
$$+ \frac{1}{2} \left[\log \left(\frac{\mathrm{T}_2}{\mathrm{T}_1} \right) - \log |\mathbf{\Sigma}_2^{-1} \mathbf{\Sigma}_1| + \sum_{j=1}^{p-1} \log \mathrm{F}_j \right] \tag{2.11}$$

where

$$\omega_{3j} = N \left(\frac{\lambda_j}{(N+1)(\lambda_j N+1)} \right)^{1/2}, \qquad \gamma_j = \left(\lambda_j + \frac{1}{N} \right)^{-1/2} \mu_j;$$

$\mathrm{T}_g \stackrel{d}{=} \chi_{N-p}^2$, $\mathrm{Z}_{gj} \stackrel{d}{=} N(0, 1)$, $g = 1, 2$, $j = 1, ..., p$; F_j *is F-distributed with* $(N-j, N-j)$ *degrees of freedom for* $j = 1, ..., p-1$; *and all random variables are mutually independent.*

If $\mathbf{y} \in \mathbf{\Pi}_2$ then

$$\hat{Q} \stackrel{d}{=} \frac{N-1}{2N} \sum_{j=1}^{p} \left[\frac{(N+1)}{\mathrm{T}_2} \mathrm{Z}_{1j}^2 - \frac{(\lambda_j^{-1} N+1)}{\mathrm{T}_1} (\tilde{\omega}_{3j} \mathrm{Z}_{1j} + (1-\tilde{\omega}_{3j}^2)^{-1/2} \mathrm{Z}_{2j} + \tilde{\gamma}_j)^2 \right]$$
$$+ \frac{1}{2} \left[\log \left(\frac{\mathrm{T}_2}{\mathrm{T}_1} \right) - \log |\mathbf{\Sigma}_2^{-1} \mathbf{\Sigma}_1| + \sum_{j=1}^{p-1} \log \mathrm{F}_j \right], \tag{2.12}$$

where, for $j = 1, ..., p$,

$$\tilde{\omega}_{3j} = N \left(\frac{1}{(N+1)(\lambda_j + N)} \right)^{1/2}, \qquad \tilde{\gamma}_j = \left(1 + \frac{\lambda_j}{N} \right)^{-1/2} \mu_j.$$

The stochastic representations in Theorem 2.1 and Corollary 2.2 imply some heuristic results about the behavior of $\hat{Q}$ as a function of the parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$. Suppose, for instance, that some μ_j in (2.2) is large, indicating that the statistical distance between the population means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ is large; and let us hold fixed the covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$. Then the corresponding γ_j in (2.4) is also large, and the right-hand side of (2.9) is dominated by the term involving γ_j^2. This results in large positive values of $\hat{Q}$ for the case in which $\mathbf{y} \in \mathbf{\Pi}_1$, with a correspondingly small estimate for the misclassification probability $P(2 \mid 1)$.

Similarly, for large values of μ_j we find that $\tilde{\gamma}_j$ in (2.7) is large, so that the right-hand side of (2.10) is dominated by the term involving $\tilde{\gamma}_j^2$. In this case we will obtain large negative values of $\hat{Q}$ for $\mathbf{y} \in \mathbf{\Pi}_2$, with a correspondingly small estimate for the misclassification probability $P(1 \mid 2)$.

Next suppose that some λ_j is large, implying that the covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are greatly dissimilar. In this situation it is more difficult to discern the general behavior of the corresponding estimates of the misclassification probabilities. To see this, consider the case in which $\mathbf{y} \in \boldsymbol{\Pi}_1$. For fixed $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and moderate sample sizes it follows from (2.3), (2.4), and (2.6) that, for large values of λ_j, ω_{3j} is close to one and γ_j and $\tilde{\gamma}_j$ both are close to zero. The resulting effect on the stochastic representation (2.9) is that $\hat{Q}$ is dominated by the term $\frac{1}{2}(\nu_{2j}-\nu_1)\, Z_{1j}^2 - \frac{1}{2}\log \lambda_j$, which is stochastically decreasing in λ_j. For increasing values of λ_j, it follows that the mean of $\hat{Q}$ attains large negative values. However the variance of this dominant term, and hence the variance of $\hat{Q}$, also increases as λ_j increases, so that that it is more difficult to determine the behavior of estimates of $P(2 \mid 1)$.

3. APPLICATIONS

In this section we apply the results of Section 2 to estimate the probabilities of misclassification for three data sets: The well-known iris data studied by R. A. Fisher [7]; a data set containing financial information about a collection of solvent and insolvent corporations; and data studied by Reaven and Miller [18] in the analysis of diabetic status through measurements on blood samples. The first two of these data sets are provided by Johnson and Wichern [8, pp. 712–716], and the third is available from Andrews and Herzberg [2, 19].

Proceeding as in our previous article [14], we provide a biplot (cf. Khattree and Naik [11]) of the data and we use the observation coordinates on the biplot to construct a 95% confidence ellipse (cf. Johnson and Wichern [8, p. 236]) corresponding to each training sample. These ellipses illustrate graphically the sample means of each training sample, the covariance structure (i.e. the shape, volume, and orientation of the sample covariance matrices) of each training sample, and provide a graphical basis for interpreting the results of our simulations.

To determine whether or not the underlying assumption of normality of $\boldsymbol{\Pi}_1$ and $\boldsymbol{\Pi}_2$ is satisfied, we apply Mardia's statistic [12] to test the hypothesis that each training sample is drawn from a multivariate normal population. Finally, we estimate the misclassification probabilities $P(1 \mid 2)$, $P(2 \mid 1)$, and TPM by performing 500,000 iterations of a Monte Carlo simulation of the stochastic representations given in Section 2. These estimates will also be compared with estimates obtained from the resubstitution and holdout methods (cf. Johnson and Wichern [8, p. 654 ff.]).

It must be noted that the stochastic representations in Section 2 are dependent upon the unknown parameters $\boldsymbol{\mu}_g$, $\boldsymbol{\Sigma}_g$, $g = 1, 2$. In estimating the misclassification probabilities $P(1 \mid 2)$, $P(2 \mid 1)$ and TPM through simulation of the stochastic representations in Theorem 2.1, we replace each unknown parameter in (2.1)–(2.2) by the value of its corresponding sample analogs as follows: For $g = 1, 2$, we replace $\boldsymbol{\mu}_g$ and $\boldsymbol{\Sigma}_g$ by the sample values of $\bar{\mathbf{y}}_g$ and $\mathbf{S}_g$, respectively; the eigenvalues $\lambda_1, \ldots, \lambda_p$ will be replaced by the sample values of $l_1, \ldots, l_p$, the eigenvalues of $\mathbf{S}_2^{-1}\mathbf{S}_1$; the orthogonal matrix $\mathbf{H}$ in (2.1) will be replaced by the sample value of an orthogonal matrix $\hat{\mathbf{H}}$ satisfying the equation

$$\mathbf{S}_2^{-1/2}\mathbf{S}_1\mathbf{S}_2^{-1/2} = \hat{\mathbf{H}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{H}}', \tag{3.1}$$

where $\hat{\boldsymbol{\Lambda}} = \mathrm{diag}(l_1, \ldots, l_p)$; and the vector $\boldsymbol{\mu}$ in (2.2) will be replaced by the sample value of

$$\hat{\boldsymbol{\mu}} = \hat{\mathbf{H}}'\mathbf{S}_2^{-1/2}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2). \tag{3.2}$$

Having replaced these parameters with their corresponding sample values, we then perform Monte Carlo simulation of the stochastic representations. As in our earlier results [14] we have developed SAS macros, written in PROC IML, for estimating by Monte Carlo methods the probabilities of misclassification for any given data set using the results in this paper; these SAS macros are available from the authors upon request.

In multivariate asymptotic distribution theory, a standard measure of the rate of convergence to a limit distribution is $N^{1/2}/p$. With this in mind we note that, in all three examples, the sample sizes are such that $N_g^{1/2}/p < 2$ for $g = 1, 2$. Hence these sample sizes are sufficiently small that asymptotic expansions of the distribution of $\hat{Q}$ cannot be expected to provide accurate results in any of our three examples, and underscores the need for exact stochastic representations in the estimation of probabilities of misclassification.

3.1. *Fisher's Iris Data*

The iris data, studied by R. A. Fisher [7] in his path-breaking work in discriminant analysis, have been widely studied; cf. [1, 8, 11, 15]. The sepal length (SEPALLEN), sepal width (SEPALWID), petal length (PETALLEN), and petal width (PETALWID) were measured, in millimeters, on 50 iris specimens of three types: *Iris setosa*, *Iris versicolor*, and *Iris virginica*. We have provided in Fig. 1 a biplot of the iris data.

From the discussions in [8, 15] and other references, it is well known that in performing classification among the three iris specimens, the most challenging problem is discrimination between the closely-located populations *Iris versicolor* and *Iris virginica*, for which we have $p = 4$

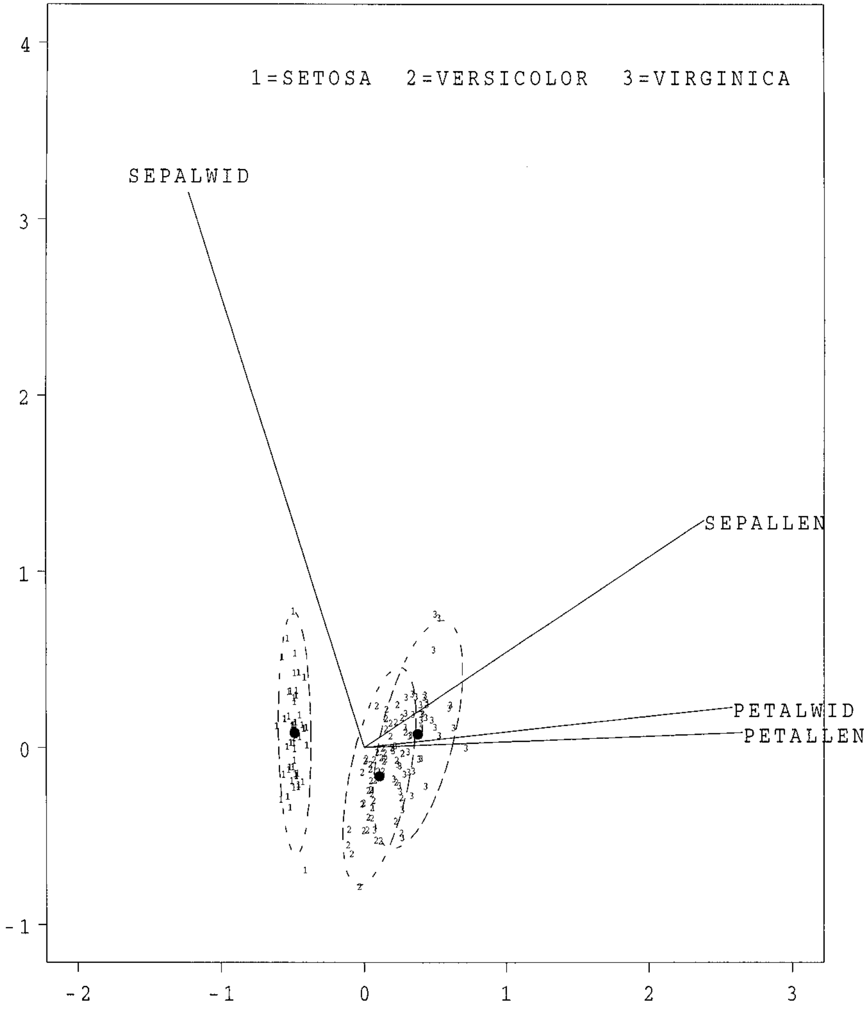

FIG. 1. Biplot of the iris data; estimated total variability explained = 95.81%.

and $N_1 = N_2 = 50$. The biplot in Fig. 1 also illustrates the relatively wide separation between *setosa* and *versicolor* or *virginica*.

We applied Mardia's test for multivariate normality to the *Iris versicolor* ($\mathbf{\Pi}_1$) and *Iris virginica* ($\mathbf{\Pi}_2$) data; at the 5% level, the test failed to reject the null hypothesis of multivariate normality in either case. The smallest P-value obtained for Mardia's skewness and kurtosis statistics, for either $\mathbf{\Pi}_1$ or $\mathbf{\Pi}_2$, was 0.0977.

Using the stochastic representation in Corollary 2.2, with all unknown parameters replaced by their sample estimates as described earlier, we

TABLE 1

Estimated Probabilities of Misclassification for the Iris Data

	Resubstitution method	Holdout method	Stochastic representation
$P(\text{versicolor} \mid \text{virginica})$	0.02	0.02	0.028924
$P(\text{virginica} \mid \text{versicolor})$	0.04	0.06	0.029400
TPM	0.03	0.04	0.029162

simulated by Monte Carlo methods the distribution of $\hat{Q}$ to estimate the probabilities of misclassification. In Table 1 we provide estimates arising from the simulation procedure. For comparative purposes we also present estimates derived through resubstitution and cross-validation procedures; thus the results in Table 1 provide a gauge of the accuracy of the resubstitution and holdout methods in the estimation of misclassification probabilities.

From Table 1, we see that the estimates of $P(1 \mid 2)$, $P(2 \mid 1)$, and TPM obtained through Monte Carlo simulation (*viz.*, 0.028924, 0.029400, and 0.029162, respectively) are, in absolute size, close to the corresponding estimates obtained through the resubstitution and holdout methods. Note that the relative differences between the Monte Carlo and the resubstitution or holdout estimates are large; however, this is to be expected given the small absolute sizes of these estimates. Although the sample size of each training data set is small, the resubstitution and holdout methods appear to perform very well here despite some well-known drawbacks (cf. [8, p. 654; 15, Chap. 10]).

The *Iris versicolor* and *Iris virginica* data were also analyzed in [13] using $\hat{Q}_1$, a discriminant criterion designed to study the situation in which the normal populations have a common mean vector (cf. [13, 14] for stochastic representations for the distribution of $\hat{Q}_1$). Based on this criterion, [13] obtained the estimates 0.2256, 0.4073, and 0.3165, for $P(1 \mid 2)$, $P(2 \mid 1)$, and TPM, respectively. Each of these estimates is significantly larger than the corresponding estimate appearing in Table 1. The larger values of the estimates obtained in [13] is due, we believe, to the governing assumption there of equality of the population means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, leading to a discriminant function depending only on the sample covariance matrices of the training data. In contrast, the present paper utilizes estimates of the parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$; so that the discriminant function $\hat{Q}$ incorporates more information from the training samples, and under fewer assumptions, than do the discriminant functions in [13] or [14]. Nevertheless, we remark that it appears difficult to obtain general results for comparison of estimates derived in [13] or [14] with estimates derived in the present paper.

3.2. *Corporate Financial Data*

Data on financial ratios of a collection of corporations are provided by Johnson and Wichern [8, pp. 712–713]. Measurements on $p = 4$ variables, CF/TD = (Cash Flow)/(Total Debt), NI/TA = (Net Income)/(Total Assets), CA/CL = (Current Assets)/(Current Liabilities), and CA/NS = (Current Assets)/(Net Sales), were collected for a group of $N_1 = 21$ insolvent and $N_2 = 25$ solvent corporations. A biplot of the data is given in Fig. 2; from the biplot and the shapes of the 95% confidence ellipses, it

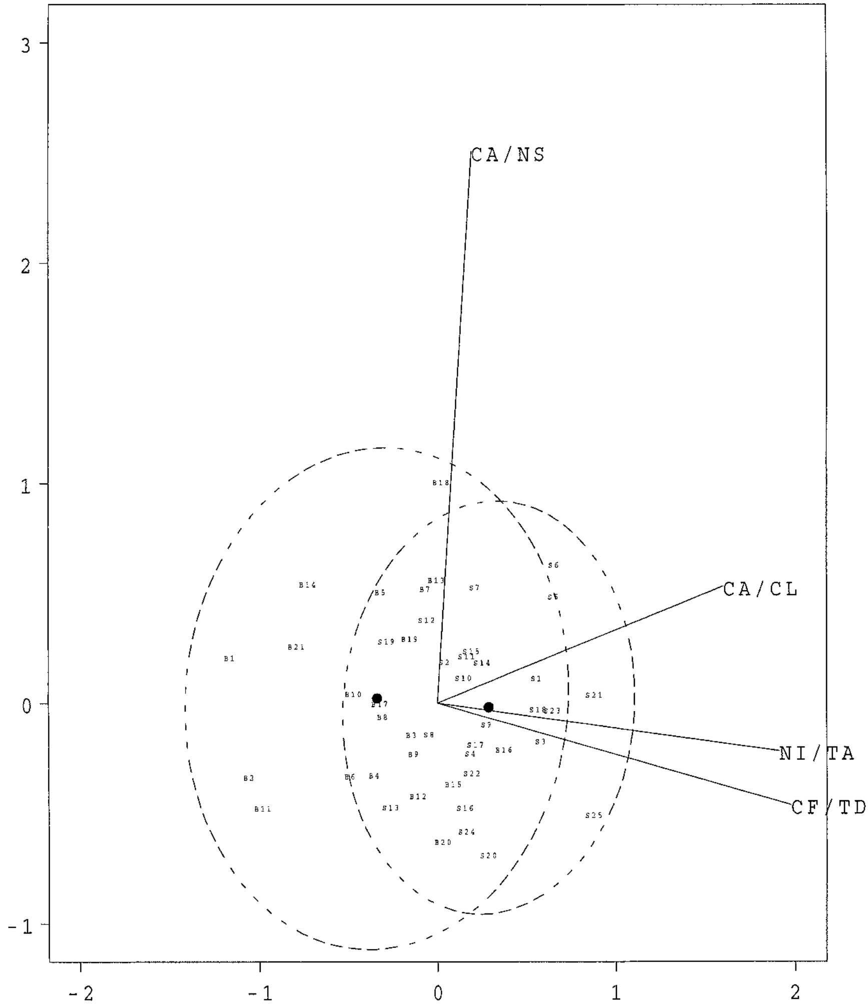

FIG. 2. Biplot of financial ratios data; estimated total variability explained = 83.11%.

TABLE 2

Estimated Probabilities of Misclassification for the Corporate Financial Data

	Resubstitution method	Holdout method	Stochastic representation
$P(\text{insolvent} \mid \text{solvent})$	0.04	0.08	0.100692
$P(\text{solvent} \mid \text{insolvent})$	0.0952	0.1429	0.099218
TPM	0.0676	0.1114	0.099955

appears that both the sample means and covariance matrices are unequal. This suggests that the quadratic discriminant function $\hat{Q}$ is appropriate for classification of these two populations.

In a test for multivariate normality, we find evidence indicating a high degree of skewness in both training samples. Mardia's skewness statistic returns P-values of 0.0001 (for the insolvent group) and 0.0006 (for the solvent group); therefore the null hypothesis of normality was rejected at the 5% level for both training samples. It is also noticeable from the biplot in Fig. 2 that the vectors labeled CF/TD and NI/TA are relatively close to each other. This indicates (cf. [11]) that the corresponding variables (denoted X_1 and X_2 in [8]) are relatively highly correlated. Nevertheless, for the purpose of direct comparison with the resubstitution and holdout classification methods, we still proceed to estimate the probabilities of misclassification using the stochastic representation for $\hat{Q}$ in Theorem 2.1.

In Table 2 we present estimates of the misclassification probabilities obtained by the resubstitution, cross-validation and Monte Carlo simulation procedures. As in the previous example, the misclassification probability estimates obtained through Monte Carlo simulation are, in absolute size, close to the corresponding estimates obtained through the resubstitution and holdout methods.

Again, there are large relative differences between estimates obtained through the resubstitution and holdout methods and the Monte Carlo simulation based on Theorem 2.1. In part, these relative differences in the estimates are due to the differing estimation methods. We also suspect that a primary reason for these relative differences in these estimates is due to the apparent non-normality of the training samples, which is counter to the normality assumptions underlying Theorem 2.1.

3.3. *Diagnosis of Diabetes through Blood Chemistry*

Reaven and Miller [18] analyzed data collected from adult subjects in a diabetes study. The purpose of the study was to investigate the relationship between measures of blood plasma glucose and insulin in a classification of subjects into three categories: normal, chemical (or sub-clinical) diabetics, or overt diabetics. We refer to [2] or [19] for the original data. After making a log-transform on two of the original variables, the variables we will use are:

1. RELWT: Relative weight (expressed as a ratio of actual weight to expected weight, given the person's height);
2. INSTEST2: The *logarithm* of plasma insulin during test (a measure of insulin response to oral glucose);
3. GLUTEST: Test-plasma glucose (a measure of glucose intolerance);
4. GLUFAST: Fasting plasma glucose; and
5. SSPG2: The *logarithm* of steady state plasma glucose response.

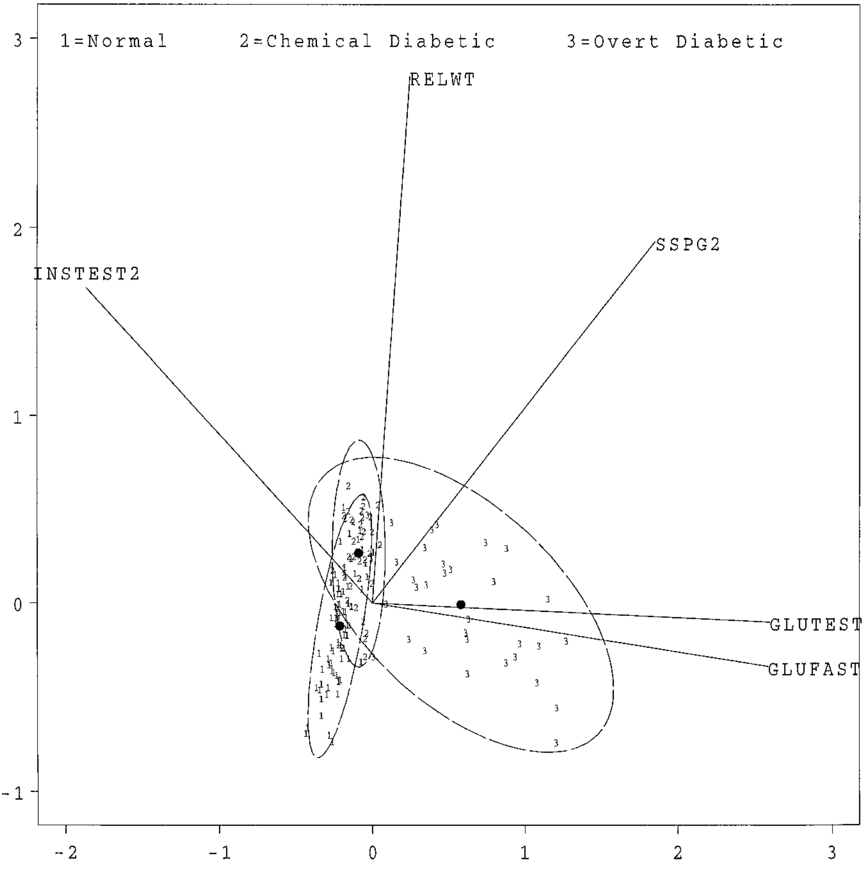

FIG. 3. Biplot of diabetes data; estimated total variability explained = 86.58%.

Figure 3 displays a biplot of the diabetes data. The biplot indicates that the Overt Diabetic group is well-separated from the remaining groups. Therefore we shall consider only the issue of discrimination between the normal and chemical diabetic groups, denoted $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$, respectively; here, we have $N_1 = 76$ and $N_2 = 36$.

The confidence ellipsoids in Fig. 3 corresponding to $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ are more elongated, and have principal axes which are less in parallel, than in previous biplots. We then may infer that the population covariance matrices corresponding to the diabetes data are more dissimilar tan in the two previous examples. At the same time, these two ellipsoids overlap to a degree greater than in the two preceding examples, and the training sample means are sufficiently close that discrimination between the two populations is a nontrivial problem.

The biplot reveals also that the majority of observations in the chemical diabetic training sample fall within the 95% confidence ellipse constructed from the training sample corresponding to the Normal group. Consequently, the biplot suggests that it is easier to misclassify an observation from $\mathbf{\Pi}_2$ into $\mathbf{\Pi}_1$ than conversely. Accordingly, we can expect that our Monte Carlo estimate of $P(2 \mid 1)$ will be relatively smaller than our estimate of $P(1 \mid 2)$.

In further exploratory analysis of the original data, we observed a high degree of skewness in two of the original variables; this explains our introduction of a log-transform on those two variables. Using Mardia's test applied to the transformed data, the null hypothesis of multivariate normality was rejected at the 5% level of significance but not at the 1% level. Specifically, the smallest P-value for the skewness or kurtosis tests for either group was calculated as 0.01930; for the test of skewness on the chemical diabetic group, the SSPG2 variable exhibited the highest tendency to skewness among all variables in either group.

Using the stochastic representation in Theorem 2.1, and with all unknown parameters estimated as described before, we performed Monte Carlo simulations to estimate the probabilities of misclassification. This

TABLE 3

Estimated Probabilities of Misclassification for the Diabetes Data

	Resubstitution method	Holdout method	Stochastic representation
P(normal | chemical diabetic)	0	0.0278	0.051358
P(chemical diabetic | normal)	0.0395	0.0526	0.037188
TPM	0.0197	0.0402	0.044273

yielded the estimates 0.051358, 0.037188, and 0.044273 for $P(1 \mid 2)$, $P(2 \mid 1)$, and TPM, respectively.

As expected, our estimate of $P(2 \mid 1)$ turns out to be relatively *smaller* than our estimate of $P(1 \mid 2)$. This should be compared with the resubstitution or holdout methods, where each estimate of $P(2 \mid 1)$ is relatively *larger* than the corresponding estimate of $P(1 \mid 2)$.

3.4. *Numerical Considerations*

The stochastic representation results in Theorem 2.1 and Corollary 2.2 are valid for all values of the parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_1$, and $\boldsymbol{\Sigma}_2$. For extreme values of these parameters, however, numerical difficulties can arise in the practical implementation of those results. To describe these difficulties, let us denote by $l_{\max}$ and $l_{\min}$ the largest and smallest eigenvalues, respectively, of the matrix $\hat{\boldsymbol{\Lambda}}$ in 3.1; then recall that the *condition number* of $\hat{\boldsymbol{\Lambda}}$, denoted by $\kappa(\hat{\boldsymbol{\Lambda}})$, is the ratio $l_{\max}/l_{\min}$.

For cases in which the matrix $\hat{\boldsymbol{\Lambda}}$ is not *well conditioned*, i.e., the condition number $\kappa(\hat{\boldsymbol{\Lambda}})$ is extremely large or small, we experienced slow convergence of the Monte Carlo simulations of the stochastic representations in Theorem 2.1 and Corollary 2.2; and conversely, if $\hat{\boldsymbol{\Lambda}}$ is well conditioned, the simulations generally converge rapidly. In the preceding examples of the Iris data and the corporate financial data, we determined values for $\kappa(\hat{\boldsymbol{\Lambda}})$ of 7.884 and 174.104, respectively; corresponding to these two greatly dissimilar values of $\kappa(\hat{\boldsymbol{\Lambda}})$, we observed rapid convergence of the Monte Carlo simulations in the case of the Iris data example and slow convergence in the case of the financial data.

Thus if the matrix $\hat{\boldsymbol{\Lambda}}$ is not well conditioned then the numerical behavior of the stochastic representations in Theorem 2.1 and Corollary 2.2 can be expected to be less than optimal. In such situations, $l_{\max}$ is substantially larger or smaller than $l_{\min}$, from which we infer that a similar result is valid for $\lambda_{\max}$ and $\lambda_{\min}$, the largest and smallest eigenvalues, respectively, of $\boldsymbol{\Lambda}$. One approach to alleviating this problem is to apply a dimension-reduction procedure, e.g., principal component analysis or biplot techniques, thereby reducing the number of variables (cf. [14, Sect. 5]). By performing dimension-reduction until the resulting matrix $\hat{\boldsymbol{\Lambda}}$ is well conditioned, we then may perform Monte Carlo simulation of the original stochastic representations applied to the reduced number of variables.

It must be kept in mind that transformations of the data may change the underlying misclassification probabilities; even in the case of data which are projected to a lower-dimensional space by principal component techniques, the classification properties of the projected data can be very different from those of the original data. In the literature, little information on the effects of transforming the data is available, and then only in the case of linear discriminant functions with large training sample sizes. For instance, the articles [4, 6] have studied the estimation of misclassification probabilities under log-normal and power transformations, respectively; they have shown that when given large data sets from closely-located log-normal populations it is unwise to use a linear discriminant function on untransformed data. More generally, the use of a linear discriminant function can lead to counterintuitive results when applied to non-normal data with large sample sizes; presumably, similar conclusions apply for cases in which the sample sizes are small. Given these results for linear discriminant functions, it is natural to expect that plug-in quadratic discriminant functions may be even more sensitive to, and more poorly behaved in the presence of, departures from normality (cf. McLachlan [15]).

4. PRELIMINARIES FOR THE PROOF OF THEOREM 2.1

Let $\mathbf{y}_1^{(1)}, \ldots, \mathbf{y}_{N_1}^{(1)}$ and $\mathbf{y}_1^{(2)}, \ldots, \mathbf{y}_{N_2}^{(2)}$ be independent random training samples from $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$, respectively, where $N_1, N_2 > p$. We denote by $\bar{\mathbf{y}}_1$ and $\bar{\mathbf{y}}_2$ the corresponding sample means and we let

$$\mathbf{A}_g = \sum_{i=1}^{N_g} (\mathbf{y}_i^{(g)} - \bar{\mathbf{y}}_g)(\mathbf{y}_i^{(g)} - \bar{\mathbf{y}}_g)' \tag{4.1}$$

denote the matrix of sums of squares and products for the gth sample. Recalling that $n_g := N_g - 1$, an unbiased estimate of the population covariance matrix $\mathbf{\Sigma}_g$ is $\mathbf{S}_g = n_g^{-1} \mathbf{A}_g$. It is well known that $\mathbf{A}_g$ has a Wishart distribution on n_g degrees of freedom, denoted $\mathbf{A}_g \stackrel{d}{=} W_p(n_g, \mathbf{\Sigma}_g)$.

Definition 4.1 (Muirhead [16, pp. 61–63]). Let $\delta \in \mathbb{C}$ such that $\mathrm{Re}(\delta) > \frac{1}{2}(p-1)$. The *multivariate gamma function* is given by

$$\Gamma_p(\delta) \equiv \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma(\delta - \tfrac{1}{2}(j-1)), \tag{4.2}$$

The following results were established in our article [14] on the equal-means plug-in discriminant functions.

Lemma 4.2 (McFarland and Richards [14]). (i) *Let $\delta \in \mathbb{C}$ such that $Re(\delta) > \frac{1}{2}(p-1)$. Then*

$$\frac{\Gamma_p(\delta)}{\Gamma(\delta - \frac{1}{2}(p-1))} = \pi^{\frac{1}{2}(p-1)} \Gamma_{p-1}(\delta). \tag{4.3}$$

(ii) *Let* $F_1, \ldots, F_{p-1}$ *be independent F-distributed random variables, where* F_j *has degrees of freedom* (N_2-j, N_1-j), $j=1, \ldots, p-1$. *Then, for* $t \in \mathbb{R}$,

$$\mathsf{E}\left[e^{it \sum_{j=1}^{p-1} \log\left(\frac{N_2-j}{N_1-j} F_j\right)}\right] = \frac{\Gamma_{p-1}(\frac{1}{2}n_1 - it)\, \Gamma_{p-1}(\frac{1}{2}n_2 + it)}{\Gamma_{p-1}(\frac{1}{2}n_1)\, \Gamma_{p-1}(\frac{1}{2}n_2)}. \tag{4.4}$$

The first part of Lemma 4.2 follows directly from Definition 4.1, while the second part follows from formulas for the characteristic function of the $\log F_j$ random variables.

We now introduce the Bessel functions of matrix argument of the second kind. We denote by $\{\mathbf{W} > \mathbf{0}\}$ the space of $p \times p$, positive-definite, symmetric matrices $\mathbf{W}$; and we denote by $d\mathbf{W}$ the Lebesgue measure on the space $\{\mathbf{W} > \mathbf{0}\}$.

Definition 4.3 (Herz [9]). Let $\mathbf{Z}$ be a $p \times p$ complex symmetric matrix and $\delta \in \mathbb{C}$. Then the *Bessel function of matrix argument of the second kind* is

$$\mathrm{B}_\delta^{(p)}(\mathbf{Z}) = \int_{\mathbf{W} > \mathbf{0}} e^{-\operatorname{tr}(\mathbf{ZW} + \mathbf{W}^{-1})}\, |\mathbf{W}|^{\delta - \frac{1}{2}(p+1)}\, d\mathbf{W}. \tag{4.5}$$

By Herz [9, p. 506 ff.], if $\operatorname{Re}(\mathbf{Z}) = \mathbf{0}$, i.e., $\mathbf{Z}$ is a purely imaginary matrix, then (4.5) is absolutely convergent if and only if $\operatorname{Re}(\delta) < -\frac{1}{2}(p-1)$. In particular, for $p=1$, in which case

$$\mathrm{B}_\delta^{(1)}(z) = \int_0^\infty e^{-zu - u^{-1}} u^{\delta - 1}\, du, \qquad z \in \mathbb{C}, \tag{4.6}$$

it follows that if $\operatorname{Re}(\delta) < 0$ then the integral (4.6) is absolutely convergent for all $z \in \mathbb{C}$ such that $\operatorname{Re}(z) = 0$.

It also follows from (4.5) that $\mathrm{B}_\delta^{(p)}(\mathbf{HZH}^{-1}) = \mathrm{B}_\delta^{(p)}(\mathbf{Z})$ for any $p \times p$ orthogonal matrix $\mathbf{H}$; therefore, if $\mathbf{Z}$ is a real or imaginary symmetric matrix then $\mathrm{B}_\delta^{(p)}(\mathbf{Z})$ depends only on the eigenvalues of $\mathbf{Z}$.

The following result is due to Herz [9, Theorem 5.10, p. 509].

Lemma 4.4 (Herz [9]). *Let* $\mathbf{Z}$ *be a* $p \times p$ *real or imaginary symmetric matrix of rank* k, *where* $k < p$, *and suppose* $\tilde{\mathbf{Z}}$ *is any* $k \times k$ *symmetric matrix whose eigenvalues are the* k *non-zero eigenvalues of* $\mathbf{Z}$. *Then*

$$B_{-\delta}^{(p)}(\mathbf{Z}) = \frac{\Gamma_p(\delta)}{\Gamma_k(\delta - \frac{1}{2}(p-k))}\, B_{-\delta + \frac{1}{2}(p-k)}^{(k)}(\tilde{\mathbf{Z}}), \tag{4.7}$$

with $\operatorname{Re}(\delta) < -\frac{1}{2}(p-k-1)$.

The connection between the Bessel functions of matrix argument and our stochastic representations begins with the following result.

LEMMA 4.5 (McFarland and Richards [14]). *Let* $\mathbf{A} \stackrel{d}{=} W_p(n, \Sigma)$, *where* $n = N-1 \geqslant p$; $\mathbf{y} \in \mathbb{R}^p$ *be a fixed vector; and* $t_1, t_2 \in \mathbb{R}$. *Then*

$$\mathsf{E}[\mathrm{e}^{i(t_1 \mathbf{y}'\mathbf{A}^{-1}\mathbf{y} + t_2 \log |\mathbf{A}|)}] = \frac{|2\Sigma|^{it_2}\, \pi^{\frac{1}{2}(p-1)} \Gamma_{p-1}(\frac{1}{2}n + it_2)}{\Gamma_p(\frac{1}{2}n)}\, B^{(1)}_{-it_2 - \frac{1}{2}(N-p)}(-\tfrac{1}{2}\, it_1 \mathbf{y}'\Sigma^{-1}\mathbf{y}). \tag{4.8}$$

The proof of Lemma 4.5 consists of noting that the left-hand side of (4.8), which can be rewritten as $\mathsf{E}[|\mathbf{A}|^{it_2}\, \mathrm{e}^{it_1 \operatorname{tr}(\mathbf{A}^{-1}\mathbf{y}\mathbf{y}')}]$, can be expressed as a constant multiple of the $p \times p$ matrix argument Bessel function $B^{(p)}_{-it_2}(-it_1 \mathbf{y}\mathbf{y}'\Sigma^{-1})$. Since the matrix $\mathbf{y}\mathbf{y}'\Sigma^{-1}$ is of rank one, then we apply (4.7) to obtain the expectation as a constant multiple of the scalar argument Bessel function $B^{(1)}_{-it_2 - \frac{1}{2}(N-p)}(-it_1 \mathbf{y}'\Sigma^{-1}\mathbf{y})$.

In the following result we list some formulas for the characteristic functions of quadratic forms in normal variables. These results have been stated in various forms in the literature; notably they can be deduced from a result given by Khatri [10, p. 446, Eq. (3.4)]. We will list the results in a format for direct application in the proof of the main result (Theorem 2.1).

LEMMA 4.6 (Khatri [10]). (i) *Let* $\mathbf{C}$ *be a real symmetric* $p \times p$ *matrix, and let* $\mathbf{v} \in \mathbb{R}^p$ *and* $\tau \in \mathbb{R}$ *be constants. If* $\mathbf{x} \stackrel{d}{=} N_p(\mathbf{0}, \Sigma)$ *then*

$$\mathsf{E}[\mathrm{e}^{it(\mathbf{x}'\mathbf{C}\mathbf{x} + \mathbf{v}'\mathbf{x} + \tau)}] = |\mathbf{I}_p - 2it\mathbf{C}\Sigma|^{-1/2} \exp(it\tau - \tfrac{1}{2}\, t^2 \mathbf{v}'\Sigma(\mathbf{I}_p - 2it\mathbf{C}\Sigma)^{-1}\, \mathbf{v}). \tag{4.9}$$

Furthermore, (4.9) *remains valid if* $\mathbf{C}$ *is a complex symmetric matrix whose imaginary part is positive-definite and* $\mathbf{v}$ *is a complex vector.*

(ii) *If* $\mathbf{x} \stackrel{d}{=} N_p(\boldsymbol{\mu}, \Sigma)$ *then*

$$\mathsf{E}[\mathrm{e}^{it\mathbf{x}'\mathbf{C}\mathbf{x}}] = |\mathbf{I}_p - 2it\mathbf{C}\Sigma|^{-1/2} \exp(it\boldsymbol{\mu}'(\mathbf{I}_p - 2it\mathbf{C}\Sigma)^{-1}\, \mathbf{C}\boldsymbol{\mu}). \tag{4.10}$$

(iii) *Let* $\gamma, \omega_1, \omega_2 \in \mathbb{R}$, $-1 < \omega_3 < 1$, *and denote*

$$G(t) := (1 + 2it\omega_1)(1 - 2it\omega_2) - 4\omega_1\omega_2\omega_3^2 t^2, \qquad t \in \mathbb{R}. \tag{4.11}$$

Let $(\mathrm{X}_1, \mathrm{X}_2)' \stackrel{d}{=} N_2(\mathbf{0}, \left(\begin{smallmatrix} 1 & \omega_3 \\ \omega_3 & 1 \end{smallmatrix}\right))$, *a bivariate normal distribution with mean vector* $\mathbf{0}$, *unit variances and correlation coefficient* ω_3. *Then*

$$\mathsf{E}[\mathrm{e}^{it(\omega_2(\mathrm{X}_2 + \gamma)^2 - \omega_1 \mathrm{X}_1^2)}] = [G(t)]^{-1/2} \exp\,(i\gamma^2\omega_2 t(1 + 2it\omega_1)[G(t)]^{-1}). \tag{4.12}$$

Proof. For the case in which the matrix $\mathbf{C}$ is real symmetric, the formula (4.9) is given by Khatri [10, p. 446, Eq. (3.4)]. Once we have established the result for real symmetric matrices C and $\mathbf{v} \in \mathbb{R}^p$, the extension to the setting in which C is a complex symmetric matrix with positive-definite imaginary part and $\mathbf{v}$ is a complex vector is obtained by analytic continuation arguments (cf. Herz [9, Sect. 2)].

Next, (4.10) is also a special case of formula (3.4) of Khatri [10].

Finally, (4.12) is the special case of (4.9) in which $p = 2$,

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \omega_3 \\ \omega_3 & 1 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} -\omega_1 & 0 \\ 0 & \omega_2 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} 0 \\ 2\gamma\omega_2 \end{pmatrix}, \quad \text{and} \quad \tau = \gamma^2\omega_2.$$

For then it follows, by performing elementary algebraic manipulations, that the left-hand sides of (4.9) and (4.12) coincide, and also that the right-hand side of (4.9) reduces to the right-hand side of (4.12). ∎

5. PROOF OF THE STOCHASTIC REPRESENTATION

First we apply an invariance argument to simplify the problem (cf. Anderson [1, p. 216–217]). Note that the statistic $\hat{Q}$ is unchanged under the action of any nonsingular affine transformation,

$$\begin{aligned} \mathbf{y}_j^{(g)*} &= \mathbf{C}(\mathbf{y}_j^{(g)} + \mathbf{v}), \qquad j = 1, \ldots, N_g,\ g = 1, 2 \\ \mathbf{y}^* &= \mathbf{C}(\mathbf{y} + \mathbf{v}), \end{aligned}$$

where $\mathbf{C}$ is a $p \times p$ nonsingular matrix and $\mathbf{v}$ is a $p \times 1$ constant vector. By choosing $\mathbf{C} = \mathbf{H}'\mathbf{\Sigma}_2^{-1/2}$ and $\mathbf{v} = -\boldsymbol{\mu}_2$ we find that no generality is lost if the parameters $(\boldsymbol{\mu}_1, \mathbf{\Sigma}_1; \boldsymbol{\mu}_2, \mathbf{\Sigma}_2)$ are replaced by $(\boldsymbol{\mu}, \mathbf{\Lambda}; \mathbf{0}, \mathbf{I}_p)$, where $\boldsymbol{\mu} = \mathbf{H}'\mathbf{\Sigma}_2^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ and $\mathbf{\Lambda} = \mathbf{H}'\mathbf{\Sigma}_2^{-1/2}\mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1/2}\mathbf{H}$, as defined earlier in (2.1). Therefore, without loss of generality, we assume that $\mathbf{\Pi}_1$ is the population $N_p(\boldsymbol{\mu}, \mathbf{\Lambda})$ and $\mathbf{\Pi}_2$ is the population $N_p(\mathbf{0}, \mathbf{I}_p)$.

Define $\mathbf{x}_g = \mathbf{y} - \bar{\mathbf{y}}_g$ for $g = 1, 2$. Then we can write the discriminant function $\hat{Q}$ in (1.3) in the form

$$\hat{Q} = -\tfrac{1}{2}(n_1 \mathbf{x}_1' \mathbf{A}_1^{-1} \mathbf{x}_1 + \log|\mathbf{A}_1|) + \tfrac{1}{2}(n_2 \mathbf{x}_2' \mathbf{A}_2^{-1} \mathbf{x}_2 + \log|\mathbf{A}_2|) + \tfrac{1}{2} p \log(n_1 n_2^{-1}), \tag{5.1}$$

where $\mathbf{A}_1 = n_1 \mathbf{S}_1 \stackrel{d}{=} W_p(n_1, \mathbf{\Lambda})$; $\mathbf{A}_2 = n_2 \mathbf{S}_2 \stackrel{d}{=} W_p(n_2, \mathbf{I}_p)$; and $(\mathbf{x}_1, \mathbf{x}_2)$, $\mathbf{A}_1$, and $\mathbf{A}_2$ are mutually independent. Then the characteristic function of $\hat{Q}$ is

$$\begin{aligned}\mathsf{E}[e^{it\hat{Q}}] &= e^{\frac{1}{2}itp\log(n_1/n_2)}\mathsf{E}_{\mathbf{x}_1,\mathbf{x}_2}\mathsf{E}_{\mathbf{A}_1}[e^{-\frac{1}{2}it(n_1\mathbf{x}_1'\mathbf{A}_1^{-1}\mathbf{x}_1+\log|\mathbf{A}_1|)}]\,\mathsf{E}_{\mathbf{A}_2}[e^{\frac{1}{2}it(n_2\mathbf{x}_2'\mathbf{A}_2^{-1}\mathbf{x}_2+\log|\mathbf{A}_2|)}]\\ &= \left(\frac{n_1^p}{n_2^p\,|\boldsymbol{\Lambda}|}\right)^{\frac{1}{2}it}\frac{\pi^{p-1}\Gamma_{p-1}(\frac{1}{2}n_1-\frac{1}{2}it)\,\Gamma_{p-1}(\frac{1}{2}n_2+\frac{1}{2}it)}{\Gamma_p(\frac{1}{2}n_1)\,\Gamma_p(\frac{1}{2}n_2)}\\ &\quad\times\mathsf{E}_{\mathbf{x}_1,\mathbf{x}_2}\left[B^{(1)}_{\frac{1}{2}it-\frac{1}{2}(N_1-p)}(\tfrac{1}{4}itn_1\mathbf{x}_1'\boldsymbol{\Lambda}^{-1}\mathbf{x}_1)\,B^{(1)}_{-\frac{1}{2}it-\frac{1}{2}(N_2-p)}(-\tfrac{1}{4}itn_2\mathbf{x}_2'\mathbf{x}_2)\right],\end{aligned} \tag{5.2}$$

where the last equality follows by applying Lemma 4.5 to evaluate the expectations with respect to $\mathbf{A}_1$ and $\mathbf{A}_2$.

Now suppose that $\mathbf{y}\in\boldsymbol{\Pi}_1$. Then the joint distribution of $(\mathbf{x}_1, \mathbf{x}_2)$ is given by

$$\begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix}=\begin{pmatrix}\mathbf{I}_p & -\mathbf{I}_p & \mathbf{O}\\ \mathbf{I}_p & \mathbf{O} & -\mathbf{I}_p\end{pmatrix}\begin{pmatrix}\mathbf{y}\\ \bar{\mathbf{y}}_1\\ \bar{\mathbf{y}}_2\end{pmatrix}\overset{d}{=}N_{2p}\left(\begin{pmatrix}\mathbf{0}\\ \boldsymbol{\mu}\end{pmatrix},\begin{bmatrix}\frac{N_1+1}{N_1}\boldsymbol{\Lambda} & \boldsymbol{\Lambda}\\ \boldsymbol{\Lambda} & \boldsymbol{\Lambda}+\frac{1}{N_2}\mathbf{I}_p\end{bmatrix}\right). \tag{5.3}$$

By Lemma 4.2, the identity $\mathsf{E}_{\mathbf{x}_1,\mathbf{x}_2}\equiv\mathsf{E}_{\mathbf{x}_1}\mathsf{E}_{\mathbf{x}_2\mid\mathbf{x}_1}$, and (5.2), we obtain

$$\begin{aligned}\mathsf{E}[e^{it\hat{Q}}] &= \mathsf{E}\left[\exp\left(\frac{1}{2}it\left\{\log\frac{n_1^p}{n_2^p\,|\boldsymbol{\Lambda}|}+\sum_{j=1}^{p-1}\log\left(\frac{N_2-j}{N_1-j}F_j\right)\right\}\right)\right]\prod_{g=1}^{2}\frac{1}{\Gamma(\frac{1}{2}(N_g-p))}\\ &\quad\times\mathsf{E}_{\mathbf{x}_1}B^{(1)}_{\frac{1}{2}it-\frac{1}{2}(N_1-p)}(\tfrac{1}{4}itn_1\mathbf{x}_1'\boldsymbol{\Lambda}^{-1}\mathbf{x}_1)\,\mathsf{E}_{\mathbf{x}_2\mid\mathbf{x}_1}B^{(1)}_{-\frac{1}{2}it-\frac{1}{2}(N_2-p)}(-\tfrac{1}{4}itn_2\mathbf{x}_2'\mathbf{x}_2),\end{aligned} \tag{5.4}$$

and it remains to simplify the expectations with respect to $\mathbf{x}_1$ and $\mathbf{x}_2$ in (5.4).

By (5.3) we have $\mathbf{x}_1\overset{d}{=}N_p(\mathbf{0}, (N_1+1)\,\boldsymbol{\Lambda}/N_1)$ and $\mathbf{x}_2\mid\mathbf{x}_1\overset{d}{=}N_p(\boldsymbol{\mu}_{2\mid1}, \boldsymbol{\Sigma}_{2\mid1})$, where

$$\boldsymbol{\mu}_{2\mid1}:=\frac{N_1}{N_1+1}\mathbf{x}_1+\boldsymbol{\mu},\qquad \boldsymbol{\Sigma}_{2\mid1}:=\frac{1}{N_1+1}\boldsymbol{\Lambda}+\frac{1}{N_2}\mathbf{I}_p. \tag{5.5}$$

Expressing the second Bessel function in (5.4) in terms of its integral representation (4.6), and interchanging integral and expectation, we obtain

$$\begin{aligned}&\mathsf{E}_{\mathbf{x}_2\mid\mathbf{x}_1}B^{(1)}_{-\frac{1}{2}it-\frac{1}{2}(N_2-p)}(-\tfrac{1}{4}itn_2\mathbf{x}_2'\mathbf{x}_2)\\ &\quad=\int_0^\infty e^{-u_2^{-1}}u_2^{-\frac{1}{2}it-\frac{1}{2}(N_2-p)-1}\mathsf{E}_{\mathbf{x}_2\mid\mathbf{x}_1}[e^{\frac{1}{4}itn_2u_2\mathbf{x}_2'\mathbf{x}_2}]\,du_2.\end{aligned} \tag{5.6}$$

To evaluate the conditional expectation in (5.6), we apply Lemma 4.6(ii); introducing the shorthand notation

$$\mathbf{M}_1 := \mathbf{I}_p - \tfrac{1}{2}\, itn_2u_2\boldsymbol{\Sigma}_{2|1} = (1 - \tfrac{1}{2}\, itn_2u_2N_2^{-1})\,\mathbf{I}_p - \tfrac{1}{2}\, itn_2u_2(N_1+1)^{-1}\,\boldsymbol{\Lambda}, \quad (5.7)$$

we obtain the result

$$\mathsf{E}_{\mathbf{x}_2|\mathbf{x}_1} B^{(1)}_{-\frac{1}{2}it-\frac{1}{2}(N_2-p)}(-\tfrac{1}{4}\, itn_2\mathbf{x}_2'\mathbf{x}_2) = \int_0^\infty \mathrm{e}^{-u_2^{-1}} u_2^{-\frac{1}{2}it-\frac{1}{2}(N_2-p)-1}\, |\mathbf{M}_1|^{-1/2} \exp[\tfrac{1}{4}\, itn_2u_2\boldsymbol{\mu}_{2|1}'\mathbf{M}_1^{-1}\boldsymbol{\mu}_{2|1}]\, du_2. \quad (5.8)$$

Now we apply the formula (5.5) for $\boldsymbol{\mu}_{2|1}$ to (5.8), substitute the resulting expression into (5.4), and interchange the integral and expectation. Then we have the result

$$\begin{aligned}
&\mathsf{E}_{\mathbf{x}_1} B^{(1)}_{\frac{1}{2}it-\frac{1}{2}(N_1-p)}\left(\frac{1}{4}\, itn_1\mathbf{x}_1'\boldsymbol{\Lambda}^{-1}\mathbf{x}_1\right) \mathsf{E}_{\mathbf{x}_2|\mathbf{x}_1} B^{(1)}_{-\frac{1}{2}it-\frac{1}{2}(N_2-p)}\left(-\frac{1}{4}\, itn_2\mathbf{x}_2'\mathbf{x}_2\right)\\
&= \int_0^\infty \mathsf{E}_{\mathbf{x}_1} B^{(1)}_{\frac{1}{2}it-\frac{1}{2}(N_1-p)}\left(\frac{1}{4}\, itn_1\mathbf{x}_1'\boldsymbol{\Lambda}^{-1}\mathbf{x}_1\right) \mathrm{e}^{-u_2^{-1}} u_2^{-\frac{1}{2}it-\frac{1}{2}(N_2-p)-1}\, |\mathbf{M}_1|^{-1/2}\\
&\quad \times \exp\left[\frac{1}{4}\, itn_2u_2\left(\frac{N_1}{N_1+1}\mathbf{x}_1+\boldsymbol{\mu}\right)' \mathbf{M}_1^{-1}\left(\frac{N_1}{N_1+1}\mathbf{x}_1+\boldsymbol{\mu}\right)\right] du_2.
\end{aligned} \quad (5.9)$$

Replacing the remaining Bessel function in (5.9) by its integral representation (4.6), collecting terms in $\mathbf{x}_1$, and interchanging integral and expectations, we obtain

$$\begin{aligned}
&\mathsf{E}_{\mathbf{x}_1} B^{(1)}_{\frac{1}{2}it-\frac{1}{2}(N_1-p)}\left(\frac{1}{4}\, itn_1\mathbf{x}_1'\boldsymbol{\Lambda}^{-1}\mathbf{x}_1\right) \mathsf{E}_{\mathbf{x}_2|\mathbf{x}_1} B^{(1)}_{-\frac{1}{2}it-\frac{1}{2}(N_2-p)}\left(-\frac{1}{4}\, itn_2\mathbf{x}_2'\mathbf{x}_2\right)\\
&= \int_0^\infty \int_0^\infty \mathsf{E}_{\mathbf{x}_1} \exp\left(\frac{1}{4}\, it\left[n_2u_2\left(\frac{N_1}{N_1+1}\mathbf{x}_1+\boldsymbol{\mu}\right)'\right.\right.\\
&\quad \left.\left.\times \mathbf{M}_1^{-1}\left(\frac{N_1}{N_1+1}\mathbf{x}_1+\boldsymbol{\mu}\right) - n_1u_1\mathbf{x}_1'\boldsymbol{\Lambda}^{-1}\mathbf{x}_1\right]\right)\\
&\quad \times |\mathbf{M}_1|^{-1/2}\, u_1^{\frac{1}{2}it-\frac{1}{2}(N_1-p)-1} u_2^{-\frac{1}{2}it-\frac{1}{2}(N_2-p)-1} \exp(-u_1^{-1}-u_2^{-1})\, du_1\, du_2.
\end{aligned} \quad (5.10)$$

Recall that $\mathbf{x}_1 \overset{d}{=} N_p(\mathbf{0}, (N_1+1)\,\boldsymbol{\Lambda}/N_1)$; in order to apply Lemma 4.6(i) to calculate the expectation in (5.10), we need to observe that the imaginary part of the complex symmetric matrix

$$n_2 u_2 \left(\frac{N_1}{N_1+1}\right)^2 \mathbf{M}_1^{-1} - n_1 u_1 \boldsymbol{\Lambda}^{-1}$$

is positive-definite. Then, by applying Lemma 4.6(i) and using the shorthand notation

$$\begin{aligned}\mathbf{M}_2 &:= \mathbf{I}_p - \frac{1}{2} it \left(\frac{N_1+1}{N_1}\right) \boldsymbol{\Lambda}^{1/2} \left[n_2 u_2 \left(\frac{N_1}{N_1+1}\right)^2 \mathbf{M}_1^{-1} - n_1 u_1 \boldsymbol{\Lambda}^{-1} \right] \boldsymbol{\Lambda}^{1/2} \\ &= \left(1 + \frac{1}{2} itn_1 u_1 \frac{N_1+1}{N_1}\right) \mathbf{I}_p - \frac{1}{2} itn_2 u_2 \frac{N_1}{N_1+1} \boldsymbol{\Lambda}^{1/2} \mathbf{M}_1^{-1} \boldsymbol{\Lambda}^{1/2}, \qquad (5.11)\end{aligned}$$

we obtain the result

$$\begin{aligned}&\mathsf{E}_{\mathbf{x}_1} \exp\left(\frac{1}{4} it \left[n_2 u_2 \left(\frac{N_1}{N_1+1}\mathbf{x}_1 + \boldsymbol{\mu}\right)' \mathbf{M}_1^{-1} \left(\frac{N_1}{N_1+1}\mathbf{x}_1 + \boldsymbol{\mu}\right) - n_1 u_1 \mathbf{x}_1' \boldsymbol{\Lambda}^{-1} \mathbf{x}_1 \right]\right) \\ &= \exp\left(\frac{1}{4} itn_2 u_2 \boldsymbol{\mu}' \mathbf{M}_1^{-1} \boldsymbol{\mu}\right) \mathsf{E}_{\mathbf{x}_1} \exp\left[\frac{1}{4} it \left\{ \mathbf{x}_1' \left[n_2 u_2 \left(\frac{N_1}{N_1+1}\right)^2 \mathbf{M}_1^{-1} - n_1 u_1 \boldsymbol{\Lambda}^{-1} \right] \mathbf{x}_1 \right.\right. \\ &\quad \left.\left. + 2 n_2 u_2 \frac{N_1}{N_1+1} \boldsymbol{\mu}' \mathbf{M}_1^{-1} \mathbf{x}_1 \right\}\right] \\ &= |\mathbf{M}_2|^{-1/2} \exp\left[\frac{1}{4} itn_2 u_2 \boldsymbol{\mu}' \mathbf{M}_1^{-1} \boldsymbol{\mu} - \frac{1}{8} t^2 n_2^2 u_2^2 \frac{N_1}{N_1+1} \boldsymbol{\mu}' \mathbf{M}_1^{-1} \boldsymbol{\Lambda}^{1/2} \mathbf{M}_2^{-1} \boldsymbol{\Lambda}^{1/2} \mathbf{M}_1^{-1} \boldsymbol{\mu}\right]. \\ &\qquad (5.12)\end{aligned}$$

After performing simple algebraic manipulations, we rewrite the right-hand side of (5.12) as

$$\begin{aligned}&|\mathbf{M}_2|^{-1/2} \exp\left[\frac{1}{4} itn_2 u_2 \boldsymbol{\mu}' \boldsymbol{\Lambda}^{-1/2} \left(\mathbf{M}_2 + \frac{1}{2} itn_2 u_2 \frac{N_1}{N_1+1} \boldsymbol{\Lambda}^{1/2} \mathbf{M}_1^{-1} \boldsymbol{\Lambda}^{1/2}\right) \right. \\ &\qquad \left. \times \mathbf{M}_2^{-1} \boldsymbol{\Lambda}^{1/2} \mathbf{M}_1^{-1} \boldsymbol{\mu}\right] \\ &\quad = |\mathbf{M}_2|^{-1/2} \exp\left[\frac{1}{4} itn_2 u_2 \left(1 + \frac{1}{2} itn_1 u_1 \frac{N_1+1}{N_1}\right) \boldsymbol{\mu}' \boldsymbol{\Lambda}^{-1/2} \mathbf{M}_2^{-1} \boldsymbol{\Lambda}^{1/2} \mathbf{M}_1^{-1} \boldsymbol{\mu}\right], \\ &\qquad (5.13)\end{aligned}$$

where the second equality follows from (5.11). Substituting (5.13) into (5.10) leads, by way of (5.4), to the result

$$\mathsf{E}[\mathrm{e}^{it\hat{Q}}] = \mathsf{E}\exp\left[\frac{1}{2}it\left\{\log\frac{n_1^p}{n_2^p\,|\boldsymbol{\Lambda}|} + \sum_{j=1}^{p-1}\log\left(\frac{N_2-j}{N_1-j}\,\mathrm{F}_j\right)\right\}\right]$$
$$\times\int_0^\infty\int_0^\infty \exp\left[\frac{1}{4}itn_2u_2\left(1+\frac{1}{2}itn_1u_1\,\frac{N_1+1}{N_1}\right)\boldsymbol{\mu}'\boldsymbol{\Lambda}^{-1/2}\mathbf{M}_2^{-1}\boldsymbol{\Lambda}^{1/2}\mathbf{M}_1^{-1}\boldsymbol{\mu}\right]$$
$$\times|\mathbf{M}_1|^{-1/2}\,|\mathbf{M}_2|^{-1/2}\,u_1^{it/2}u_2^{-it/2}\prod_{g=1}^{2}\frac{u_g^{-\frac{1}{2}(N_g-p)-1}\mathrm{e}^{-u_g^{-1}}}{\Gamma(\frac{1}{2}(N_g-p))}\,du_g. \tag{5.14}$$

From (5.11) we deduce that

$$|\mathbf{M}_2| = \left|\left(1+\frac{1}{2}itn_1u_1\,\frac{N_1+1}{N_1}\right)\mathbf{I}_p - \frac{1}{2}itn_2u_2\,\frac{N_1}{N_1+1}\,\boldsymbol{\Lambda}^{1/2}\mathbf{M}_1^{-1}\boldsymbol{\Lambda}^{1/2}\right|$$
$$= \left|\left(1+\frac{1}{2}itn_1u_1\,\frac{N_1+1}{N_1}\right)\mathbf{I}_p - \frac{1}{2}itn_2u_2\,\frac{N_1}{N_1+1}\,\mathbf{M}_1^{-1}\boldsymbol{\Lambda}\right|;$$

hence, by (5.7),

$$|\mathbf{M}_1|\,|\mathbf{M}_2|$$
$$= \left|\left(1+\frac{1}{2}itn_1u_1\,\frac{N_1+1}{N_1}\right)\mathbf{M}_1 - \frac{1}{2}itn_2u_2\,\frac{N_1}{N_1+1}\,\boldsymbol{\Lambda}\right|$$
$$\equiv \left|\left(1+\frac{1}{2}itn_1u_1\,\frac{N_1+1}{N_1}\right)\left(\mathbf{I}_p - \frac{1}{2}itn_2u_2\left(\boldsymbol{\Lambda}+\frac{1}{N_2}\mathbf{I}_p\right)\right) - \frac{1}{4}t^2n_1n_2u_1u_2\boldsymbol{\Lambda}\right|$$
$$= \prod_{j=1}^{p}\left[\left(1+\frac{1}{2}itn_1u_1\,\frac{N_1+1}{N_1}\right)\left(1-\frac{1}{2}itn_2u_2\left(\lambda_j+\frac{1}{N_2}\right)\right) - \frac{1}{4}n_1n_2u_1u_2\lambda_jt^2\right]. \tag{5.15}$$

Applying (5.11) we find that

$$\mathbf{M}_1\boldsymbol{\Lambda}^{-1/2}\mathbf{M}_2\boldsymbol{\Lambda}^{1/2}$$
$$= \mathbf{M}_1\boldsymbol{\Lambda}^{-1/2}\left[\left(1+\frac{1}{2}itn_1u_1\,\frac{N_1+1}{N_1}\right)\mathbf{I}_p - \frac{1}{2}itn_2u_2\,\frac{N_1}{N_1+1}\,\boldsymbol{\Lambda}^{1/2}\mathbf{M}_1^{-1}\boldsymbol{\Lambda}^{1/2}\right]\boldsymbol{\Lambda}^{1/2}$$
$$= \left(1+\frac{1}{2}itn_1u_1\,\frac{N_1+1}{N_1}\right)\mathbf{M}_1 - \frac{1}{2}itn_2u_2\,\frac{N_1}{N_1+1}\,\boldsymbol{\Lambda}$$
$$= \left(1+\frac{1}{2}itn_1u_1\,\frac{N_1+1}{N_1}\right)\left(\mathbf{I}_p - \frac{1}{2}itn_2u_2\left(\boldsymbol{\Lambda}+\frac{1}{N_2}\mathbf{I}_p\right)\right) - \frac{1}{4}t^2n_1n_2u_1u_2\boldsymbol{\Lambda}. \tag{5.16}$$

Then

$$\begin{aligned}
&\boldsymbol{\mu}'\boldsymbol{\Lambda}^{-1/2}\mathbf{M}_2^{-1}\boldsymbol{\Lambda}^{1/2}\mathbf{M}_1^{-1}\boldsymbol{\mu}\\
&\quad=\boldsymbol{\mu}'(\mathbf{M}_1\boldsymbol{\Lambda}^{-1/2}\mathbf{M}_2\boldsymbol{\Lambda}^{1/2})^{-1}\,\boldsymbol{\mu}\\
&\quad=\boldsymbol{\mu}'\left[\left(1+\frac{1}{2}\,itn_1u_1\,\frac{N_1+1}{N_1}\right)\left(\mathbf{I}_p-\frac{1}{2}\,itn_2u_2\left(\boldsymbol{\Lambda}+\frac{1}{N_2}\,\mathbf{I}_p\right)\right)-\frac{1}{4}\,t^2n_1n_2u_1u_2\boldsymbol{\Lambda}\right]^{-1}\boldsymbol{\mu}\\
&\quad=\sum_{j=1}^{p}\mu_j^2\left[\left(1+\frac{1}{2}\,itn_1u_1\,\frac{N_1+1}{N_1}\right)\left(1-\frac{1}{2}\,itn_2u_2\left(\lambda_j+\frac{1}{N_2}\right)\right)-\frac{1}{4}\,t^2n_1n_2u_1u_2\lambda_j\right]^{-1}.
\end{aligned}\tag{5.17}$$

Combining (5.14), (5.15), and (5.16) we obtain

$$\begin{aligned}
\mathsf{E}[\mathrm{e}^{it\hat{Q}}]&=\mathsf{E}\exp\left[\frac{1}{2}\,it\left\{\log\frac{n_1^p}{n_2^p\,|\boldsymbol{\Lambda}|}+\sum_{j=1}^{p-1}\log\left(\frac{N_2-j}{N_1-j}\,\mathrm{F}_j\right)\right\}\right]\\
&\quad\times\int_0^\infty\int_0^\infty\exp\left[\frac{1}{4}\,itn_2u_2\left(1+\frac{1}{2}\,itn_1u_1\,\frac{N_1+1}{N_1}\right)\right.\\
&\quad\times\sum_{j=1}^{p}\mu_j^2\left[\left(1+\frac{1}{2}\,itn_1u_1\,\frac{N_1+1}{N_1}\right)\left(1-\frac{1}{2}\,itn_2u_2\left(\lambda_j+\frac{1}{N_2}\right)\right)\right.\\
&\quad\left.\left.-\frac{1}{4}\,t^2n_1n_2u_1u_2\lambda_j\right]^{-1}\right]\\
&\quad\times\prod_{j=1}^{p}\left[\left(1+\frac{1}{2}\,itn_1u_1\,\frac{N_1+1}{N_1}\right)\left(1-\frac{1}{2}\,itn_2u_2\left(\lambda_j+\frac{1}{N_2}\right)\right)\right.\\
&\quad\left.-\frac{1}{4}\,n_1n_2u_1u_2\lambda_jt^2\right]^{-1/2}u_1^{it/2}u_2^{-it/2}\prod_{g=1}^{2}\frac{u_g^{-\frac{1}{2}(N_g-p)-1}\mathrm{e}^{-u_g^{-1}}}{\Gamma(\frac{1}{2}(N_g-p))}\,du_g.
\end{aligned}\tag{5.18}$$

For $j=1,\ldots,p$, define

$$\omega_1=\frac{1}{4}\,n_1u_1\,\frac{N_1+1}{N_1},\qquad\omega_{2j}=\frac{1}{4}\,n_2u_2\left(\lambda_j+\frac{1}{N_2}\right),\tag{5.19}$$

and set

$$G_j(t):=(1+2\omega_1it)(1-2\omega_{2j}it)-4\omega_1\omega_{2j}\omega_{3j}^2t^2,$$

where ω_{3j} is defined in (2.3). Recalling the definition of γ_j given in (2.4), we find that (5.18) becomes

$$\mathsf{E}[\mathrm{e}^{it\hat{Q}}] = \int_0^\infty \int_0^\infty \mathsf{E}[\mathrm{e}^{\frac{1}{2}it\left\{\log \frac{n_1^p u_1}{n_2^p |\Lambda| u_2} + \sum_{j=1}^{p-1} \log\left(\frac{N_2-j}{N_1-j}\mathrm{F}_j\right)\right\}}]$$
$$\times \prod_{j=1}^{p} [G_j(t)]^{-1/2} \exp(i\gamma_j^2 \omega_{2j} t(1+2\omega_1 it)[G_j(t)]^{-1})$$
$$\times \prod_{g=1}^{2} \frac{u_g^{-\frac{1}{2}(N_g-p)-1} \mathrm{e}^{-u_g^{-1}}}{\Gamma(\frac{1}{2}(N_g-p))}\, du_g. \tag{5.20}$$

By Lemma 4.6(iii), (5.20) becomes

$$\mathsf{E}[\mathrm{e}^{it\hat{Q}}] = \int_0^\infty \int_0^\infty \mathsf{E}[\mathrm{e}^{\frac{1}{2}it\left\{\log \frac{n_1^p u_1}{n_2^p |\Lambda| u_2} + \sum_{j=1}^{p-1} \log\left(\frac{N_2-j}{N_1-j}\mathrm{F}_j\right)\right\}}]$$
$$\times \mathsf{E}[\mathrm{e}^{it\sum_{j=1}^{p}\{\omega_{2j}(\mathrm{X}_{2j}+\gamma_j)^2 - \omega_1 \mathrm{X}_{1j}^2\}}] \prod_{g=1}^{2} \frac{u_g^{-\frac{1}{2}(N_g-p)-1} \mathrm{e}^{-u_g^{-1}}}{\Gamma(\frac{1}{2}(N_g-p))}\, du_g, \tag{5.21}$$

where $(\mathrm{X}_{1j}, \mathrm{X}_{2j})' \overset{d}{=} N_2(\mathbf{0}, \begin{pmatrix} 1 & \omega_{3j} \\ \omega_{3j} & 1 \end{pmatrix})$, $j = 1, \ldots, p$.

On making the transformation $t_1 = 2u_1^{-1}$ and $t_2 = 2u_2^{-1}$, which has the Jacobian $4t_1^{-2}t_2^{-2}$, we obtain (5.21) in the form

$$\mathsf{E}[\mathrm{e}^{it\hat{Q}}] = \int_0^\infty \int_0^\infty g_{N_1-p}(t_1)\, g_{N_2-p}(t_2)\, \mathsf{E}[\mathrm{e}^{it\mathrm{W}(t_1, t_2)}]\, dt_1\, dt_2, \tag{5.22}$$

where g_k is the density function of a chi-squared random variable with k degrees of freedom;

$$\mathrm{W}(t_1, t_2) \overset{d}{=} \frac{1}{2} \sum_{j=1}^{p} [\nu_{2j}(\mathrm{X}_{2j}+\gamma_j)^2 - \nu_1 \mathrm{X}_{1j}^2]$$
$$+ \frac{1}{2}\log\left(\frac{n_1^p t_2}{n_2^p t_1}\right) - \frac{1}{2}\log|\Lambda| + \frac{1}{2}\sum_{j=1}^{p-1} \log\left(\frac{N_2-j}{N_1-j}\mathrm{F}_j\right); \tag{5.23}$$

ν_1 and ν_{2j} are the random variables defined in (2.5); and γ_j is defined in (2.4).

Since each $(\mathrm{X}_{1j}, \mathrm{X}_{2j})' \overset{d}{=} N_2(\mathbf{0}, \begin{pmatrix} 1 & \omega_{3j} \\ \omega_{3j} & 1 \end{pmatrix})$ then it is well known that

$$(\mathrm{X}_{1j}, \mathrm{X}_{2j}) \overset{d}{=} (\mathrm{Z}_{1j}, \omega_{3j}\mathrm{Z}_{1j} + (1-\omega_{3j}^2)^{1/2}\,\mathrm{Z}_{2j}), \tag{5.24}$$

where the Z_{gj} are independent and identically distributed standard normal variables. Substituting (5.24) into (5.23), and interchanging expectation and integrals in (5.22), we obtain

$$\mathsf{E}[e^{it\hat{Q}}] = \mathsf{E} \int_0^\infty \int_0^\infty g_{N_1-p}(t_1)\, g_{N_2-p}(t_2)\, e^{it\mathrm{W}(t_1,\, t_2)}\, dt_1\, dt_2,$$
$$\equiv \mathsf{E}[e^{it\mathrm{W}}] \tag{5.25}$$

where W is the random variable given in (2.9). Therefore we conclude from (5.25) that $\hat{Q} \stackrel{d}{=} \mathrm{W}$.

Next suppose that $\mathbf{y} \in \mathbf{\Pi}_2$. Bearing in mind the affine transformation applied at the outset of the proof, we maintain the assumption that $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ are the populations $N_p(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ and $N_p(\mathbf{0}, \mathbf{I}_p)$, respectively. Then the joint distribution of $(\mathbf{x}_1, \mathbf{x}_2)$ is given by

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_p & -\mathbf{I}_p & \mathbf{O} \\ \mathbf{I}_p & \mathbf{O} & -\mathbf{I}_p \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \bar{\mathbf{y}}_1 \\ \bar{\mathbf{y}}_2 \end{pmatrix} \stackrel{d}{=} N_{2p}\left(\begin{pmatrix} -\boldsymbol{\mu} \\ \mathbf{0} \end{pmatrix}, \begin{bmatrix} \mathbf{I}_p + \dfrac{1}{N_1}\boldsymbol{\Lambda} & \mathbf{I}_p \\ \mathbf{I}_p & \dfrac{N_2+1}{N_2}\mathbf{I}_p \end{bmatrix} \right). \tag{5.26}$$

By Lemma 4.2, the identity $\mathsf{E}_{\mathbf{x}_1, \mathbf{x}_2} \equiv \mathsf{E}_{\mathbf{x}_2} \mathsf{E}_{\mathbf{x}_1 \mid \mathbf{x}_2}$, and (5.2), we obtain

$$\begin{aligned} \mathsf{E}[e^{it\hat{Q}}] &= \mathsf{E}\left[\exp\left(\frac{1}{2}\, it \left\{ \log \frac{n_1^p}{n_2^p\, |\boldsymbol{\Lambda}|} + \sum_{j=1}^{p-1} \log\left(\frac{N_2 - j}{N_1 - j}\, \mathrm{F}_j \right) \right\} \right) \right] \prod_{g=1}^{2} \frac{1}{\Gamma(\frac{1}{2}(N_g - p))} \\ &\quad \times \mathsf{E}_{\mathbf{x}_2} B^{(1)}_{-\frac{1}{2}it - \frac{1}{2}(N_2-p)}\left(-\frac{1}{4}\, itn_2 \mathbf{x}_2' \mathbf{x}_2 \right) \mathsf{E}_{\mathbf{x}_1 \mid \mathbf{x}_2} B^{(1)}_{\frac{1}{2}it - \frac{1}{2}(N_1-p)} \\ &\quad \times \left(\frac{1}{4}\, itn_1 \mathbf{x}_1' \boldsymbol{\Lambda}^{-1} \mathbf{x}_1 \right), \end{aligned} \tag{5.27}$$

and it remains to simplify the expectations with respect to $\mathbf{x}_1$ and $\mathbf{x}_2$ in (5.4).

By (5.26) we have $\mathbf{x}_2 \stackrel{d}{=} N_p(\mathbf{0}, (N_2+1)\, \mathbf{I}_p / N_2)$ and $\mathbf{x}_1 \mid \mathbf{x}_2 \stackrel{d}{=} N_p(\boldsymbol{\mu}_{1\mid 2}, \boldsymbol{\Sigma}_{1\mid 2})$, where

$$\boldsymbol{\mu}_{1\mid 2} := \frac{N_2}{N_2+1}\, \mathbf{x}_2 - \boldsymbol{\mu}, \qquad \boldsymbol{\Sigma}_{1\mid 2} := \frac{1}{N_2+1}\, \mathbf{I}_p + \frac{1}{N_1}\, \boldsymbol{\Lambda}. \tag{5.28}$$

Expressing the second Bessel function in (5.27) in terms of its integral representation (4.6), and interchanging integral and expectation, we obtain

$$\begin{aligned} &\mathsf{E}_{\mathbf{x}_1 \mid \mathbf{x}_2} B^{(1)}_{\frac{1}{2}it - \frac{1}{2}(N_1-p)}(\tfrac{1}{4}\, itn_1 \mathbf{x}_1' \boldsymbol{\Lambda}^{-1} \mathbf{x}_1) \\ &\quad = \int_0^\infty e^{-u_1^{-1}}\, u_1^{\frac{1}{2}it - \frac{1}{2}(N_1-p) - 1}\, \mathsf{E}_{\mathbf{x}_1 \mid \mathbf{x}_2} [e^{-\frac{1}{4} itn_1 u_1 \mathbf{x}_1' \boldsymbol{\Lambda}^{-1} \mathbf{x}_1}]\, du_1. \end{aligned} \tag{5.29}$$

To evaluate the conditional expectation in (5.29), we apply Lemma 4.6(ii); introducing the shorthand notation

$$\mathbf{M}_3 := \mathbf{I}_p + \tfrac{1}{2}\, itn_1 u_1 \boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}_{1|2} = (1 + \tfrac{1}{2}\, itn_1 u_1 N_1^{-1})\, \mathbf{I}_p + \tfrac{1}{2}\, itn_1 u_1 (N_2+1)^{-1}\, \boldsymbol{\Lambda}^{-1}, \tag{5.30}$$

we obtain the result

$$\begin{aligned} &\mathsf{E}_{\mathbf{x}_1|\mathbf{x}_2} B^{(1)}_{\frac{1}{2}it - \frac{1}{2}(N_1-p)}(\tfrac{1}{4}\, itn_1 \mathbf{x}_1' \boldsymbol{\Lambda}^{-1}\mathbf{x}_1) \\ &\quad = \int_0^\infty \mathrm{e}^{-u_1^{-1}} u_1^{\frac{1}{2}it - \frac{1}{2}(N_1-p)-1}\, |\mathbf{M}_3|^{-1/2} \exp[-\tfrac{1}{4}\, itn_1 u_1 \boldsymbol{\mu}_{1|2}' (\boldsymbol{\Lambda}\mathbf{M}_3)^{-1}\, \boldsymbol{\mu}_{1|2}]\, du_1. \end{aligned} \tag{5.31}$$

Now we apply the formula (5.28) for $\boldsymbol{\mu}_{1|2}$ to (5.31), substitute the resulting expression into (5.27), and interchange the integral and expectation. Then we have the result

$$\begin{aligned} &\mathsf{E}_{\mathbf{x}_2} B^{(1)}_{-\frac{1}{2}it - \frac{1}{2}(N_2-p)} \left(-\frac{1}{4}\, itn_2 \mathbf{x}_2'\mathbf{x}_2\right) \mathsf{E}_{\mathbf{x}_1|\mathbf{x}_2} B^{(1)}_{\frac{1}{2}it - \frac{1}{2}(N_1-p)} \left(\frac{1}{4}\, itn_1 \mathbf{x}_1' \boldsymbol{\Lambda}^{-1}\mathbf{x}_1\right) \\ &\quad = \int_0^\infty \mathsf{E}_{\mathbf{x}_2} B^{(1)}_{-\frac{1}{2}it - \frac{1}{2}(N_2-p)} \left(-\frac{1}{4}\, itn_2 \mathbf{x}_2'\mathbf{x}_2\right) \mathrm{e}^{-u_1^{-1}} u_1^{\frac{1}{2}it - \frac{1}{2}(N_1-p)-1}\, |\mathbf{M}_3|^{-1/2} \\ &\qquad \times \exp\left[-\frac{1}{4}\, itn_1 u_1 \left(\frac{N_2}{N_2+1}\,\mathbf{x}_2 - \boldsymbol{\mu}\right)' (\boldsymbol{\Lambda}\mathbf{M}_3)^{-1} \left(\frac{N_2}{N_2+1}\,\mathbf{x}_2 - \boldsymbol{\mu}\right)\right] du_1 \end{aligned} \tag{5.32}$$

Replacing the remaining Bessel function in (5.32) by its integral representation (4.6), collecting terms in $\mathbf{x}_2$, and interchanging integral and expectations, we obtain

$$\begin{aligned} &\mathsf{E}_{\mathbf{x}_2} B^{(1)}_{-\frac{1}{2}it - \frac{1}{2}(N_2-p)} \left(-\frac{1}{4}\, itn_2 \mathbf{x}_2'\mathbf{x}_2\right) \mathsf{E}_{\mathbf{x}_1|\mathbf{x}_2} B^{(1)}_{\frac{1}{2}it - \frac{1}{2}(N_1-p)} \left(\frac{1}{4}\, itn_1 \mathbf{x}_1' \boldsymbol{\Lambda}^{-1}\mathbf{x}_1\right) \\ &\quad = \int_0^\infty \int_0^\infty \mathsf{E}_{\mathbf{x}_2} \exp\left(-\frac{1}{4}\, it \left[n_1 u_1 \left(\frac{N_2}{N_2+1}\,\mathbf{x}_2 - \boldsymbol{\mu}\right)' (\boldsymbol{\Lambda}\mathbf{M}_3)^{-1}\right.\right. \\ &\qquad \left.\left. \times \left(\frac{N_2}{N_2+1}\,\mathbf{x}_2 - \boldsymbol{\mu}\right) - n_2 u_2 \mathbf{x}_2'\mathbf{x}_2 \right]\right) \\ &\qquad \times |\mathbf{M}_3|^{-1/2}\, u_1^{\frac{1}{2}it - \frac{1}{2}(N_1-p)-1}\, u_2^{-\frac{1}{2}it - \frac{1}{2}(N_2-p)-1} \exp(-u_1^{-1} - u_2^{-1})\, du_1\, du_2. \end{aligned} \tag{5.33}$$

Recall that $\mathbf{x}_2 \stackrel{d}{=} N_p(\mathbf{0}, (N_2+1)\,\mathbf{I}_p/N_2)$; in order to apply Lemma 4.6(i) to calculate the expectation in (5.10), we need to observe that the imaginary part of the complex symmetric matrix

$$-n_1 u_1 \left(\frac{N_2}{N_2+1}\right)^2 (\mathbf{\Lambda M}_3)^{-1} + n_2 u_2 \mathbf{I}_p$$

is positive-definite. Then, by applying Lemma 4.6(i) and using the short-hand notation

$$\begin{aligned}
\mathbf{M}_4 &:= \mathbf{I}_p - \frac{1}{2}\, it\, \frac{N_2+1}{N_2} \left(n_2 u_2 \mathbf{I}_p - n_1 u_1 \left(\frac{N_2}{N_2+1}\right)^2 (\mathbf{\Lambda M}_3)^{-1} \right) \\
&= \left(1 - \frac{1}{2}\, it n_2 u_2\, \frac{N_2+1}{N_2}\right) \mathbf{I}_p + \frac{1}{2}\, it n_1 u_1\, \frac{N_2}{N_2+1} (\mathbf{\Lambda M}_3)^{-1},
\end{aligned} \tag{5.34}$$

we obtain the result

$$\begin{aligned}
&\mathsf{E}_{\mathbf{x}_2} \exp\left(-\frac{1}{4}\, it \left[n_1 u_1 \left(\frac{N_2}{N_2+1}\, \mathbf{x}_2 - \boldsymbol{\mu} \right)' (\mathbf{\Lambda M}_3)^{-1} \left(\frac{N_2}{N_2+1}\, \mathbf{x}_2 - \boldsymbol{\mu} \right) - n_2 u_2 \mathbf{x}_2' \mathbf{x}_2 \right] \right) \\
&\quad = \exp\left(-\frac{1}{4}\, it n_1 u_1 \boldsymbol{\mu}' (\mathbf{\Lambda M}_3)^{-1} \boldsymbol{\mu} \right) \\
&\qquad \times \mathsf{E}_{\mathbf{x}_2} \exp\left[\frac{1}{4}\, it \left\{ \mathbf{x}_2' \left\{ n_2 u_2 \mathbf{I}_p - n_1 u_1 \left(\frac{N_2}{N_2+1} \right)^2 (\mathbf{\Lambda M}_3)^{-1} \right\} \mathbf{x}_2 \right. \right. \\
&\qquad \left. \left. + 2 n_1 u_1\, \frac{N_2}{N_2+1}\, \boldsymbol{\mu}' (\mathbf{\Lambda M}_3)^{-1} \mathbf{x}_2 \right\} \right] \\
&\quad = |\mathbf{M}_4|^{-1/2} \exp\left[-\frac{1}{4}\, it n_1 u_1 \boldsymbol{\mu}' (\mathbf{\Lambda M}_3)^{-1} \boldsymbol{\mu} \right. \\
&\qquad \left. - \frac{1}{8}\, t^2 n_1^2 u_1^2\, \frac{N_2}{N_2+1}\, \boldsymbol{\mu}' (\mathbf{\Lambda M}_3)^{-1} \mathbf{M}_4^{-1} (\mathbf{\Lambda M}_3)^{-1} \boldsymbol{\mu} \right].
\end{aligned} \tag{5.35}$$

By performing simple algebraic manipulations, we rewrite the right-hand side of (5.35) as

$$\begin{aligned}
&|\mathbf{M}_4|^{-1/2} \exp\left[-\frac{1}{4}\, it n_1 u_1 \boldsymbol{\mu}' (\mathbf{\Lambda M}_3)^{-1} \mathbf{M}_4^{-1} \left(\mathbf{M}_4 - \frac{1}{2}\, it n_1 u_1\, \frac{N_2}{N_2+1} (\mathbf{\Lambda M}_3)^{-1} \right) \boldsymbol{\mu} \right] \\
&\quad = |\mathbf{M}_4|^{-1/2} \exp\left[-\frac{1}{4}\, it n_1 u_1 \left(1 - \frac{1}{2}\, it n_2 u_2\, \frac{N_2+1}{N_2} \right) \boldsymbol{\mu}' (\mathbf{\Lambda M}_3)^{-1} \mathbf{M}_4^{-1} \boldsymbol{\mu} \right],
\end{aligned} \tag{5.36}$$

where the second equality follows from (5.34). Substituting (5.36) into (5.33) leads, by way of (5.27), to the result

$$\mathsf{E}[\mathrm{e}^{it\hat{Q}}] = \mathsf{E}\exp\left[\frac{1}{2}it\left\{\log\frac{n_1^p}{n_2^p\,|\boldsymbol{\Lambda}|} + \sum_{j=1}^{p-1}\log\left(\frac{N_2-j}{N_1-j}\,\mathrm{F}_j\right)\right\}\right]$$
$$\times\int_0^\infty\int_0^\infty \exp\left[-\frac{1}{4}itn_1u_1\left(1-\frac{1}{2}itn_2u_2\,\frac{N_2+1}{N_2}\right)\boldsymbol{\mu}'(\boldsymbol{\Lambda}\mathbf{M}_3)^{-1}\,\mathbf{M}_4^{-1}\boldsymbol{\mu}\right]$$
$$\times|\mathbf{M}_3|^{-1/2}\,|\mathbf{M}_4|^{-1/2}\,u_1^{it/2}u_2^{-it/2}\prod_{g=1}^{2}\frac{u_g^{-\frac{1}{2}(N_g-p)-1}\mathrm{e}^{-u_g^{-1}}}{\Gamma(\frac{1}{2}(N_g-p))}\,du_g. \tag{5.37}$$

From (5.34) and simple algebraic manipulations, we deduce that

$$\mathbf{M}_3\mathbf{M}_4 = \left(1-\frac{1}{2}itn_2u_2\,\frac{N_2+1}{N_2}\right)\mathbf{M}_3 + \frac{1}{2}itn_1u_1\,\frac{N_2}{N_2+1}\,\boldsymbol{\Lambda}^{-1}\,|$$
$$= \left(1-\frac{1}{2}itn_2u_2\,\frac{N_2+1}{N_2}\right)\left(\mathbf{I}_p+\frac{1}{2}itn_1u_1\left(\frac{1}{N_1}\,\mathbf{I}_p+\boldsymbol{\Lambda}^{-1}\right)\right)-\frac{1}{4}t^2n_1n_2u_1u_2\boldsymbol{\Lambda}^{-1};$$

hence, by (5.30),

$$|\mathbf{M}_3|\,|\mathbf{M}_4| = |\mathbf{M}_3\mathbf{M}_4|$$
$$= \prod_{j=1}^{p}\left[\left(1+\frac{1}{2}itn_1u_1(\lambda_j^{-1}+N_1^{-1})\right)\left(1-\frac{1}{2}itn_2u_2\,\frac{N_2+1}{N_2}\right)\right.$$
$$\left.-\frac{1}{4}n_1n_2u_1u_2\lambda_j^{-1}t^2\right]. \tag{5.38}$$

Since the matrices $\mathbf{M}_3$, $\mathbf{M}_4$, and $\boldsymbol{\Lambda}$ are diagonal and hence are commutative, we also have

$$\boldsymbol{\mu}'(\boldsymbol{\Lambda}\mathbf{M}_3)^{-1}\,\mathbf{M}_4^{-1}\boldsymbol{\mu}$$
$$= \boldsymbol{\mu}'\boldsymbol{\Lambda}^{-1}(\mathbf{M}_3\mathbf{M}_4)^{-1}\,\boldsymbol{\mu}$$
$$= \sum_{j=1}^{p}\mu_j^2\lambda_j^{-1}\left[\left(1+\frac{1}{2}itn_1u_1(\lambda_j^{-1}+N_1^{-1})\right)\left(1-\frac{1}{2}itn_2u_2\,\frac{N_2+1}{N_2}\right)\right.$$
$$\left.-\frac{1}{4}n_1n_2u_1u_2\lambda_j^{-1}\right]^{-1}. \tag{5.39}$$

Combining (5.37), (5.38), and (5.39) we obtain

$$\begin{aligned}
\mathsf{E}[e^{it\hat{Q}}] = \mathsf{E}\exp\left[\frac{1}{2}it\left\{\log\frac{n_1^p}{n_2^p\,|\Lambda|}+\sum_{j=1}^{p-1}\log\left(\frac{N_2-j}{N_1-j}\,\mathrm{F}_j\right)\right\}\right] \\
\times\int_0^\infty\int_0^\infty \exp\left[-\frac{1}{4}itn_1u_1\left(1-\frac{1}{2}itn_2u_2\,\frac{N_2+1}{N_2}\right)\right. \\
\times\sum_{j=1}^{p}\mu_j^2\lambda_j^{-1}\left[\left(1+\frac{1}{2}itn_1u_1(\lambda_j^{-1}+N_1^{-1})\right)\left(1-\frac{1}{2}itn_2u_2\,\frac{N_2+1}{N_2}\right)\right. \\
\left.\left.-\frac{1}{4}n_1n_2u_1u_2\lambda_j^{-1}t^2\right]^{-1}\right] \\
\times\prod_{j=1}^{p}\left[\left(1+\frac{1}{2}itn_1u_1(\lambda_j^{-1}+N_1^{-1})\right)\left(1-\frac{1}{2}itn_2u_2\,\frac{N_2+1}{N_2}\right)\right. \\
\left.-\frac{1}{4}n_1n_2u_1u_2\lambda_j^{-1}t^2\right]^{-1/2} \\
\times u_1^{it/2}u_2^{-it/2}\prod_{g=1}^{2}\frac{u_g^{-\frac{1}{2}(N_g-p)-1}e^{-u_g^{-1}}}{\Gamma(\frac{1}{2}(N_g-p))}\,du_g. \qquad (5.40)
\end{aligned}$$

For $j=1, ..., p$, define

$$\tilde{\omega}_1=-\frac{1}{4}n_2u_2\,\frac{N_2+1}{N_2}, \qquad \tilde{\omega}_{2j}=-\frac{1}{4}n_1u_1(\lambda_j^{-1}+N_1^{-1}) \qquad (5.41)$$

and set

$$\tilde{G}_j(t):=(1+2\tilde{\omega}_1it)(1-2\tilde{\omega}_{2j}it)-4\tilde{\omega}_1\tilde{\omega}_{2j}\tilde{\omega}_{3j}^2t^2,$$

where $\tilde{\omega}_{3j}$ is defined in (2.6). Recalling the definition of $\tilde{\gamma}_j$ given in (2.7), we deduce that (5.39) can be rewritten as

$$\begin{aligned}
\mathsf{E}[e^{it\hat{Q}}] = \int_0^\infty\int_0^\infty \mathsf{E}[e^{\frac{1}{2}it\left\{\log\frac{n_1^pu_1}{n_2^p\,|\Lambda|\,u_2}+\sum_{j=1}^{p-1}\log\left(\frac{N_2-j}{N_1-j}\mathrm{F}_j\right)\right\}}] \\
\times\prod_{j=1}^{p}[\tilde{G}_j(t)]^{-1/2}\exp(i\tilde{\gamma}_j^2\tilde{\omega}_{2j}t(1+2\tilde{\omega}_1it)[\tilde{G}_j(t)]^{-1}) \\
\times\prod_{g=1}^{2}\frac{u_g^{-\frac{1}{2}(N_g-p)-1}e^{-u_g^{-1}}}{\Gamma(\frac{1}{2}(N_g-p))}\,du_g. \qquad (5.42)
\end{aligned}$$

By Lemma 4.6(iii), (5.41) becomes

$$\mathsf{E}[\mathrm{e}^{it\hat{Q}}] = \int_0^\infty \int_0^\infty \mathsf{E}[\mathrm{e}^{\frac{1}{2}it\left\{\log \frac{n_1^p u_1}{n_2^p |\Lambda| u_2} + \sum_{j=1}^{p-1} \log\left(\frac{N_2-j}{N_1-j}\mathrm{F}_j\right)\right\}}]$$

$$\times \mathsf{E}[\mathrm{e}^{-it\sum_{j=1}^{p}\{\tilde{\omega}_{2j}(\tilde{\mathrm{X}}_{2j}+\tilde{\gamma}_j)^2 - \tilde{\omega}_1 \tilde{\mathrm{X}}_{1j}^2\}}] \prod_{g=1}^{2} \frac{u_g^{-\frac{1}{2}(N_g-p)-1} \mathrm{e}^{-u_g^{-1}}}{\Gamma(\frac{1}{2}(N_g-p))} \, du_g, \qquad (5.43)$$

where $(\tilde{\mathrm{X}}_{1j}, \tilde{\mathrm{X}}_{2j})' \stackrel{d}{=} N_2(\mathbf{0}, \left(\begin{smallmatrix} 1 & \tilde{\omega}_{3j} \\ \tilde{\omega}_{3j} & 1 \end{smallmatrix}\right))$, $j = 1, \ldots, p$.

On making the transformation $t_1 = 2u_1^{-1}$ and $t_2 = 2u_2^{-1}$, which has the Jacobian $4t_1^{-2}t_2^{-2}$, we obtain (5.42) in the form

$$\mathsf{E}[\mathrm{e}^{it\hat{Q}}] = \int_0^\infty \int_0^\infty g_{N_1-p}(t_1)\, g_{N_2-p}(t_2)\, \mathsf{E}[\mathrm{e}^{it\tilde{\mathrm{W}}(t_1,t_2)}] \, dt_1 \, dt_2, \qquad (5.44)$$

where g_k is the density function of a chi-squared random variable with k degrees of freedom;

$$\tilde{\mathrm{W}}(t_1, t_2) \stackrel{d}{=} \frac{1}{2} \sum_{j=1}^{p} [\tilde{\nu}_1 \tilde{\mathrm{X}}_{1j}^2 - \tilde{\nu}_{2j}(\tilde{\mathrm{X}}_{2j} + \tilde{\gamma}_j)^2]$$

$$+ \frac{1}{2}\log\left(\frac{n_1^p t_2}{n_2^p t_1}\right) - \frac{1}{2}\log|\Lambda| + \frac{1}{2}\sum_{j=1}^{p-1} \log\left(\frac{N_2-j}{N_1-j}\mathrm{F}_j\right); \qquad (5.45)$$

$\tilde{\nu}_1$ and $\tilde{\nu}_{2j}$ are defined in (2.8); and $\tilde{\gamma}_j$ is defined in (2.7).

Since each $(\tilde{\mathrm{X}}_{1j}, \tilde{\mathrm{X}}_{2j})' \stackrel{d}{=} N_2(\mathbf{0}, \left(\begin{smallmatrix} 1 & \tilde{\omega}_{3j} \\ \tilde{\omega}_{3j} & 1 \end{smallmatrix}\right))$ then it is well-known that

$$(\tilde{\mathrm{X}}_{1j}, \tilde{\mathrm{X}}_{2j}) \stackrel{d}{=} (\mathrm{Z}_{1j}, \tilde{\omega}_{3j}\mathrm{Z}_{1j} + (1-\tilde{\omega}_{3j}^2)^{1/2}\,\mathrm{Z}_{2j}), \qquad (5.46)$$

where the Z_{gj} are independent and identically distributed standard normal variables. Substituting (5.46) into (5.45), and interchanging expectation and integrals in (5.44), we obtain

$$\mathsf{E}[\mathrm{e}^{it\hat{Q}}] = \mathsf{E}\int_0^\infty \int_0^\infty g_{N_1-p}(t_1)\, g_{N_2-p}(t_2)\, \mathrm{e}^{it\tilde{\mathrm{W}}(t_1,t_2)} \, dt_1 \, dt_2,$$

$$\equiv \mathsf{E}[\mathrm{e}^{it\tilde{\mathrm{W}}}], \qquad (5.47)$$

where $\tilde{\mathrm{W}}$ is the random variable given in (2.10). Therefore we conclude from (5.47) that $\hat{Q} \stackrel{d}{=} \tilde{\mathrm{W}}$.

ACKNOWLEDGMENTS

We thank the referees for their invaluable comments on this article.

REFERENCES

1. T. W. Anderson, “An Introduction to Multivariate Statistical Analysis,” 2nd ed., Wiley, New York, 1984.
2. D. F. Andrews and A. M. Herzberg, “Data: A Collection of Problems from Many Fields for the Student and Research Worker,” Springer-Verlag, New York, 1985.
3. M. S. Bartlett and N. W. Please, Discrimination in the case of zero mean differences, *Biometrika* **50** (1963), 17–21.
4. J. J. Beauchamp and D. S. Robson, Transformation considerations in discriminant analysis, *Comm. Statist. Simulation Comput.* **15** (1986), 147–179.
5. A. H. Bowker, A representation of Hotelling’s T^2 and Anderson’s classification statistic W in terms of simple statistics, *in* “Studies in Item Analysis and Prediction” (H. Solomon, Ed.), pp. 285–292, Stanford Univ. Press, Stanford, CA, 1961.
6. E. F. Chingánda and K. Subrahmaniam, Robustness of the linear discriminant function to nonnormality: Johnson’s system, *J. Statist. Plann. Inference* **3** (1979), 69–77.
7. R. A. Fisher, The use of multiple measurements in taxonomic problems, *Ann. Eugen.* **7** (1936), pp. 179–188.
8. R. A. Johnson and D. W. Wichern, “Applied Multivariate Statistical Analysis,” 4th ed., Prentice Hall, 1998.
9. C. S. Herz, Bessel functions of matrix argument, *Ann. of Math.* **61** (1955), 474–522.
10. C. G. Khatri, Quadratic forms in normal variables, *in* “Handbook of Statistics” (P. R. Krishnaiah, Ed.), Vol. 1, pp. 443–469, North-Holland, New York, 1980.
11. R. Khattree and D. N. Naik, “Applied Multivariate Statistics with SAS® Software,” SAS Institute, Cary, NC, 1996.
12. K. V. Mardia, Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies, *Sankhyā Ser. B* **36** (1974), 115–128.
13. H. R. McFarland, III, “The Exact Distributions of “Plug-In” Discriminant Functions in Multivariate Analysis,” doctoral dissertation, University of Virginia, 1998.
14. H. R. McFarland, III and D. St. P. Richards, Exact misclassification probabilities for plug-in normal quadratic discriminant functions. I. The equal-means case, *J. Multivariate Anal.* **77** (2001), 21–53.
15. G. J. McLachlan, “Discriminant Analysis and Statistical Pattern Recognition,” Wiley, New York, 1992.
16. R. J. Muirhead, “Aspects of Multivariate Statistical Theory,” Wiley, New York, 1982.
17. M. Okamoto, Discrimination for variance matrices, *Osaka J. Math.* **13** (1961), 1–39.
18. G. M. Reaven and R. G. Miller, An attempt to define the nature of chemical diabetes using multidimensional analysis, *Diabetologia* **16** (1979), 17–24.
19. SAS Institute, “Logistic Regression Examples Using the SAS® System,” SAS Institute, Cary, NC, 1995.
20. P. Stocks, A biometric investigation of twins, Part II, *Ann. Eugen.* **5** (1933), 1–55.